

A Robust Reputation System based on Rating Separation and Trustworthiness of Ratings

Abstract

e-마켓플레이스에서 제공하는 판매자의 평판은 해당 판매자가 수행한 모든 거래에 대한 구매자들의 평점을 종합하여 (aggregation) 측정된다. 이렇게 구해진 판매자의 평판은 예비 소비자들이 판매자에 대한 신용을 가늠하는데 활용된다. 그러나 평점 기반의 판매자 평판은, (1) 만약 구매자가 평판조작 등의 악의적 목적으로 불공평한 평점 (unfair rating) 매길 경우, (2) 평판 계산시 구매자의 평점에 판매자의 행동과 상품의 품질에 대한 평가가 혼재되어있음을 고려하지 못하는 경우, 왜곡될 수 있다. 본 논문은 이 두 가지 문제를 해결하여 예비 소비자들에게 신뢰할 수 있는 판매자 평판을 제공하는 것을 목표로 한다.

The overall buyers' satisfaction with a seller, which we call *seller reputation*, is quantified as the aggregated score of all ratings given to the seller. Seller reputation is used as a key factor to figure out the credibility of the seller. There is, however, no guarantee of the reliability of seller reputation. The reputation of a seller could be distorted (1) when some buyers intentionally give unfair ratings to the seller and/or (2) when the reputation system does not capture the fact that a buyer's rating is the compound score of *the capability of a seller and the quality of an item*. In this paper, we propose a robust reputation system that provides reliable seller reputation by reducing the influence of unfair ratings and by separating *the capability of a seller and the quality of an item being sold*.

Keywords

Trust, Rating, Rating Separation, Unfair rating, Attacker, Reputation System, e-Marketplace

1. Introduction

e-마켓플레이스 (e.g., eBay)는 다수의 소비자와 다수의 판매자가 인터넷 공간에서 상품을 거래할 수 있도록 결제, 배송 등의 다양한 서비스를 지원하는 온라인 사이트이다 [1]. 판매자를 직접 만나 판매자의 신용을 판단할 수 있는 오프라인의 거래와는 달리 온라인에서는 소비자가 판매자를 직접 만날 수 없기 때문에 상품 거래 전에 판매자의 신용을 정확하게 파악하는 것은 매우 어렵다 [2-6].

e-마켓플레이스에서 상품을 구매한 소비자 (구매자¹)들은 평점을 통해 판매자와의 거래 (구

¹ 본 논문에서는 물건을 구매한 소비자를 구매자로 정의하되 문맥에 따라 소비자와 구매자를

체적으로, 판매자의 행동과 상품의 품질)에 대해 평가를 내린다. 판매자에 대한 구매자 전반의 만족도를 판매자의 평판 (reputation)이라 한다. 판매자의 평판은 해당 판매자가 수행한 모든 거래에 대한 구매자들의 평점을 종합하여 (aggregation) 측정한다. e-마켓플레이스는 평판시스템 (reputation system)을 구축, 판매자의 평판을 제공함으로써 소비자들이 신뢰할 수 있는 판매자와 거래할 수 있도록 돕는다 [7-15].

E-marketplaces, such as eBay.com, enable a large number of buyers and sellers to do transactions online [1]. Compared to offline, online buyers have more difficult time to identify reliable sellers with credibility, since they cannot meet sellers face to face [2-6]. To help potential buyers find reliable sellers, most e-marketplaces offer a system that provides seller reputation [7-15]. In e-marketplace, the buyer expresses his level of satisfaction with his transaction through an online rating. The overall buyers' satisfaction with a seller, called seller reputation, is quantified as the aggregated score of all ratings given to the seller.

판매자 평판은 구매자들의 집단지성 (독립적인 다양한 구매자들의 협력의 산출물)의 결과이다 [22-24]. 만약 (1) 다수의 구매자들이 정직하게 평가에 참여하고 (2) 평판시스템이 판매자의 행동에 대한 평가만을 사용하여 평판을 계산한다면 그 값의 신뢰가 높아 질 수 있다. 그러나 이와 반대 상황으로 (1) 구매자가 평판조작 등의 악의적 목적으로 불공평한 평점 (unfair rating) 매기며 (2) 평판시스템이 *구매자의 평점은 판매자의 행동과 상품의 품질에 대한 평가가 혼재되어있음*을 고려하지 못한다면 판매자의 평판은 왜곡될 수 있다.

신뢰할 수 있는 평판시스템은 첫째, 평가가 혼재된 평점에서 판매자에 대한 평가를 분리하여 평판을 계산해야 하며 둘째, 조작된 평점을 매겨 판매자의 평판을 왜곡하는 구매자들의 불공평한 평점 공격 (unfair rating attacks)에 견고해야 한다. 저자들은 이 두 가지 문제를 해결하기 위한 새로운 방법들을 개발했다.

The seller reputation will be more reliable (1) when buyers participate in evaluation with honesty and (2) when it is based on the buyers' satisfaction² about the capability of the seller (and not about the quality of the item being sold). If some buyers intentionally give unfair ratings to a seller or if the reputation system does not take into account the fact that a buyer's rating is the compound score of the capability of a seller and the quality of an item, the seller reputation could be distorted. A reliable reputation system should be able to separate the buyer's rating

interchangeably하게 사용한다.

² Note that the buyer's satisfaction is judged based on two aspects of a transaction: the capability of the seller and the quality of the item being sold.

into a score about the seller and a score about the item and use only the seller score to compute the seller reputation. It should also be robust against unfair rating attacks by the buyers who try to distort the seller reputation. In this paper, we propose a new method to solve these two problems.

일반적으로 판매자의 평판은 구매자 평점들의 산술평균으로 계산된다. 구매자의 평가 기준은 크게 판매자의 행동에 대한 평가와 상품의 품질에 대한 평가로 분류될 수 있다. 본 논문에서는 두 가지 평가가 혼재된 구매자 평점을 판매자 점수와 상품 점수로 분리하여 판매자 점수만을 이용한 평판 즉, 판매자 행동에 대한 평판을 계산하며, 이 방법을 평점 분리 (RATING SEPARATION: RS)라고 부른다. 판매자 행동에 대한 평판은 예비 소비자들이 빠른 배송과 친절 한 서비스를 제공하는 판매자를 선택할 수 있도록 돕는다.

The most common way to compute the seller reputation is to use the average of ratings. The rating, the buyer's evaluation for a transaction between a seller and a buyer, is based on two criteria. The first criterion is the capability of the seller, such as the safe and prompt delivery and the promptness of the replies to buyer's questions. The second criterion is the quality of an item. In this paper, we separate the buyer's rating into 'seller score' and 'item score' and compute the seller reputation using the seller score only. This method is called *Rating Separation (RS)*. The reputation based on seller scores would aid potential buyers in their search for reliable sellers who provide fast delivery and better service.

e-마켓플레이스에는 예비 소비자들에게 거짓된 정보를 제공하려는 목적을 가지고 부정 행위 (cheating behavior)를 저지르는 구매자들이 존재한다. 예를 들면, 판매자는 자신의 평판을 인위적으로 높이기 위해 구매자를 매수, 매수 당한 구매자는 공모 판매자에게는 의도적으로 높은 평점을 또한 경쟁 판매자³에게는 의도적으로 낮은 평점을 매겨 거짓된 정보를 양산한다. 본 논문에서는 거짓된 정보 생산을 목적으로 평점을 매기는 구매자를 unfair rating attacker (UR-attacker)라고 하고, 이들의 부정행위를 unfair rating attack (UR-attack)이라고 한다. UR-attack 처리를 위한 다양한 방법 [6, 7, 9, 11]들이 소개되었지만, 기존의 방법들은 각 방법에서 처리 가능한 UR-attack의 종류가 한정적이라는 한계를 지닌다. 본 논문에서는 기존의 소개된 UR-attack들에 견고한 평판 계산을 수행하며, 이 방법을 BETTER SELLER DISCOVERY (BSD)라고 부른다. BSD는 모든 소비자 평점에 대한 신뢰도를 측정한 후 판매자의 평판을 계산할 때 평점의 신뢰도에 따라 가중치를 다르게 부여하는 평판 계산 방법이다. BSD를 수행하면 UR-attackers의 조

³ 경쟁 판매자는 매수 당한 소비자와 공모하는 공모 판매자와 동일한 또는 유사한 상품을 판매하고 있는 잠재적인 경쟁자를 의미한다.

작된 평점들의 영향력을 줄여 판매자의 평판을 보정할 수 있게 된다.

In e-marketplace, there exist various types of cheating behaviors [7-18]. To raise their reputations, for example, some sellers may collude with a group of buyers who would provide unfairly positive ratings to them while giving unfairly negative ratings to the competing sellers. We call the buyers who supply false information *unfair rating attackers* (*UR-attackers*) and call their cheating behaviors *unfair rating attacks* (*UR-attacks*). Several methods for handling UR-attacks have been proposed [6, 7, 9, 11], but each method deals with only a subset of UR-attacks. In this paper, we propose a method called *Better Seller Discovery* (*BSD*) that is robust against any types of UR-attacks under consideration. BSD evaluates the level of trustworthiness of an individual rating and re-computes the seller reputation by assigning different weights to the individual rating according to rating trustworthiness. By taking into account the trustworthiness of each rating when computing the seller reputation, BSD can reduce the influence of UR-attackers.

최종적으로 본 논문에서는 평점 분리 기반의 평판 계산 방법인 RS와 평점 신뢰도 측정 기반의 평판 보정 방법인 BSD를 접목시킨 판매자 평판 계산 방법 RS&BSD를 제안한다. 제안하는 방법은 BSD를 통해 unfair ratings 영향력을 줄인 후 RS를 사용, 판매자 점수만을 이용하여 판매자의 평판을 측정하는 방법이다.

We propose a robust reputation system, called RS&BSD, that integrates both RS and BSD. This reputation system reduces the influence of UR-attackers through BSD and computes the seller reputation using only the seller score separated by RS.

Furthermore, we have developed a simulation model that reflects realistic e-marketplace situations. In existing studies [7-15, 19-21], e-marketplace is simulated based on simplified buyer-seller behavior patterns. It is commonly assumed that a buyer evaluates a transaction based on the capability of a seller only. In real e-marketplace situations, however, a buyer's evaluation of a transaction is based on two aspects of the transaction: the capability of a seller and the quality of an item. Therefore, our simulation model captures the tripartite relationships between buyers, sellers, and items.

The paper is organized as follows: Section 2 explains why a reputation system needs rating separation. Section 3 describes rating separation (RS), better seller discovery (BSD), and RS&BSD, the integrated version of RS and BSD. Section 4 verifies the robustness of our reputation system under comprehensive UR-attacks through extensive experiments in comparisons with the state-of-the-art approaches. Section 5 concludes our paper.

2. Necessity of rating separation

그림 1은 실제 상황에서 판매자에 대한 평점과 상품의 품질에 대한 평점의 분리가 필요한 이유를 보여준다. 그림 1-1의 소비자는 검은색 모자를 판매하는 판매자와의 거래를 평가할 때 판매자의 행동에 대한 평가와 검은색 모자의 품질에 대한 평가를 모두 고려하여 평점을 산정한다. 소비자는 두 가지 평가기준으로 거래를 평가하지만 평판 시스템에서는 하나의 평점만을 산정할 수 있다. 그림 1-2에서 소비자들은 검은색 모자를 판매하는 판매자와의 거래에 각각 3, 4, 2의 평점을 매겼고, 산술평균으로 평판을 도출할 경우 판매자의 평판은 3점으로 결정된다. 소비자가 매긴 각 평점에는 판매자의 행동에 대한 평가와 상품의 품질에 대한 평가가 모두 담겨있다. 오직 판매자의 행동에 대한 평판을 계산하기 위해서는 그림 1-3처럼 평점 분리를 통해 구매자 평점에서 판매자 점수를 추출해야 한다.

In this section, we explain why the real-world reputation system needs rating separation. Figure 1 shows an example of an existing reputation system. In Figure 1-1, the buyer has entered the rating of 3 on the transaction with the seller who provides a black cap. Note that the buyer assigns a single rating to the transaction even though his evaluation of the transaction is based on two criteria (the seller's capability and the quality of the item). In Figure 1-2, three buyers have assigned different ratings of 3, 4, and 2, respectively, to their transactions with the seller. The average of ratings is typically used when computing seller reputations, and as a result, the seller reputation is calculated as 3.

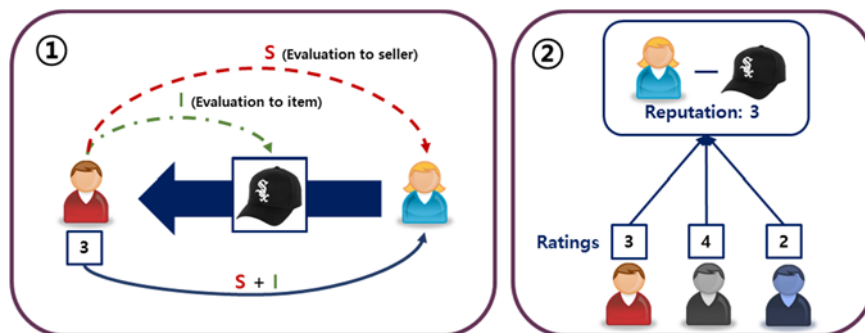


Fig. 1: Real-world reputation system

기존의 평판 시스템들은 소비자들이 두 가지 평가 기준을 바탕으로 평점을 산정한다는 것을 고려하지 않은 채로 단순히 평점들을 종합 (aggregation)해서 판매자 평판 계산에 사용한다. 일반적으로 판매자의 평판은 평점들의 산술평균으로 계산된다. 이 경우, 평판의 근원이 판매자의 행동인지 상품의 품질인지 알 수가 없기 때문에 오히려 예비 소비자들에게 혼란을 야기할 수 있다. (이건 왜 번역하지 않았나요?)

Existing reputation systems simply aggregate ratings and use them to calculate seller ratings

without considering that consumers will be calculating ratings based on two criteria. In general, the seller's reputation is calculated as the arithmetic average of the ratings. In this case, it may be confusing to prospective consumers, because it is not clear whether the origin of the reputation is the seller or the quality of the goods.

As shown in Figures 1-1 & 1-2, existing reputation systems calculates the seller reputation as the average of the buyers' ratings. In existing reputation systems, therefore, it is impossible to discern whether the seller reputation is attributable to the seller's capability or the item's quality, thereby causing confusion to potential buyers.

그림이 독자의 이해를 돕는 것 같지 않습니다. 1-3은 일부러 없앤 건가요???

예를 들면, 판매자의 평판이 높게 형성된 상황이 (1) 판매자가 배송을 지체했음에도 불구하고 상품의 품질이 우수해서 높은 평점을 받는 경우 또는 (2) 상품의 품질이 생각보다 떨어졌지만 판매자의 빠른 배송과 친절한 서비스로 인해 높은 평점을 받는 경우라면 예비 소비자들은 판매자에 대한 소비자들의 높은 평점의 원인이 '판매자의 행동'인지 '상품의 품질'인지 명확하게 알 수 없다. 또 다른 예로, 판매자의 평판이 낮게 형성된 상황이 (3) 빠른 배송과 친절한 서비스에도 불구하고 상품의 품질이 맘에 들지 않아서 낮은 평점을 매기는 경우 또는 (4) 상품의 품질은 맘에 들지만 판매자의 불친절한 태도 때문에 낮은 평점을 매기는 경우에도 또한 예비 소비자들은 판매자에 대한 소비자들의 낮은 평점의 원인이 '판매자의 행동'인지 '상품의 품질'인지 명확하게 알 수 없다.

For example, the transaction may receive a high rating for several reasons: (1) the item is of an excellent quality despite of late delivery, or (2) the seller has provided detailed reply and fast delivery even when the item's quality does not meet the buyer's expectation. In both cases, using existing reputation systems, it is impossible to distinguish the cause of a high rating. Similarly, it is difficult to tell whether a transaction receives a low rating because of the seller's poor performance or the item's poor quality.

(만일 "제품에 대한 확실한 평가가 소비자에게 존재하는 경우에는" 이라는 의미가 포함되어야 할 것 같습니다. 온라인 상품 구매 시 소비자들이 가장 우려하는 점이 바로 판매자의 행동이다. 만약 판매자의 행동 즉, 구매한 물건이 빠르고 안전하게 도착하는지, 물건에 하자가 있을 때 손쉽게 환불이 가능한지 등을 수치화하여 파악 할 수 있으면 예비 소비자들은 보다 안전하게 거래에 응할 수 있다. 따라서 구매자의 평점에서 판매자의 점수를 추출, 이 점수만을 이용하여 판매자 평판을 계산하는 평점 분리가 필요하다. (왜 문단을 정확하게 번역하지 않았나요???)

It is the seller's behavior that buyers are most concerned about when purchasing online

products. If the buyer can determine whether the purchased item would arrive quickly and safely, or if the item can be easily refunded if there is a defect, the buyer would feel more safe with the transaction. Therefore, it is necessary to separate the seller score and the item score, and to calculate the seller's reputation using only this score. (seller's score를 정의한 적이 있나요???) Abstract 에서는 언급했지만 본문에는 아직 정의되지 않은 것 같습니다.)

Buyers would be more confident in online transactions when provided with the information about the seller's capability. We claim that a reliable reputation system needs to separate the rating between the capability of a seller and the quality of an item. The separation of the two evaluation criteria will provide more precise information about the seller's capability.

3. 제안하는 방법 Proposed method

저자들은 판매자 평점을 구매자에 대한 점수와 상품에 대한 점수를 분리하여 평판을 계산하는 RS와 평점의 신뢰도를 측정하여 평판을 보정하는 BSD를 접목시킨 판매자 평판 계산 방법 RS&BSD을 제안한다.

We propose RS&BSD, a robust reputation system that computes the seller reputation using only seller scores (through RS) and adjusts the seller reputation by taking into account the level of trustworthiness of ratings (through BSD).

3.1 평점 분리 방안 Rating Separation

Rating Separation (RS) consists of three parts: *initial Seller Rating Separation (Initial-SRS)*, *Seller Rating Separation (SRS)*, and *Item Rating Separation (IRS)*. Initial-SRS is performed only once at the beginning. It is followed by IRS, and then SRS and IRS are repeated until seller reputations converge. The overall algorithm of RS is presented below.

Algorithm 1 Rating Separation (RS)

Input: e-marketplace data; set of buyers, set of sellers, set of items, set of ratings

Output: reputations of all sellers e^*

//INITIAL SELLER RATING SEPARATION (Initial-SRS)

Build a set of initial seller clusters (C^s)

for each cluster

for each seller s

Compute initial seller reputation e from set of ratings

Compute seller score e'

for each seller s

Computer seller reputation e''

Compute normalized seller reputation e^* from e''

//ITEM RATING SEPARATION (IRS)

Build a set of item clusters (C^m) based on all the e^* s

```

for each cluster
  for each item  $m$ 
    Compute item reputation  $z$  from set of ratings
    Compute item score  $z'$ 
  for each item  $m$ 
    Computer item reputation  $z''$ 
    Compute normalized item reputation  $z^*$  from  $z''$ 
repeat
  //SELLER RATING SEPARATION (SRS)
  Make  $C^s$  based on all the  $z^*$ s
  for each cluster
    for each seller  $s$ 
      Compute seller reputation  $e$  from set of ratings
      Compute seller score  $e'$ 
    for each seller  $s$ 
      Computer seller reputation  $e''$ 
      Compute normalized seller reputation  $e^*$  from  $e''$ 
  //ITEM RATING SEPARATION (IRS)
  Make  $C^m$  based on all the  $e^*$ s
  for each cluster
    for each item  $m$ 
      Compute item reputation  $z$  from set of ratings
      Compute item score  $z'$ 
    for each item  $m$ 
      Computer item reputation  $z''$ 
      Compute normalized item reputation  $z^*$  from  $z''$ 
until convergence of  $e^*$  and  $z^*$ 

```

When computing the seller score in Initial-SRS, we assume the buyer's evaluations on the quality of the same item tend to be similar. That is, if two buyers have bought same items from two different sellers and have given different ratings to their transactions, two sellers have different capabilities. Under this assumption, we aggregate the sellers with the same item into clusters, compute the relative scores among the sellers in the same cluster and designate them as (initial) seller scores.

Note that the item score can be computed in a similar manner. We call this computation process that isolates item scores to compute item reputations *Item Rating Separation (IRS)*. Compared to Initial-SRS that clusters sellers based on same items, IRS groups items based on the initial seller score attained by Initial-SRS. Item scores are computed as the relative scores among the sellers in the same cluster.

The RS process starts with Initial-SRS. Using the seller reputations obtained from Initial SRS, IRS computes item reputations. Then, SRS re-computes seller reputations, and IRS re-computes item reputations. Because SRS changes seller scores, and the changed seller scores may change item scores, and so on, we apply SRS and IRS iteratively until all seller reputations converge. We

describe the detail of each computation process in the following.

소비자의 평점에서 판매자 점수를 분리하는 방법은 다음과 같다. 첫째, 동일 상품을 기준으로 판매자들을 클러스터링 (clustering) 한다. 그림 2의 예제에는 총 다섯 명의 판매자 ($s_1 \sim s_5$)와 다섯 개의 상품 ($m_1 \sim m_5$)이 존재한다. 초기 판매자 클러스터링을 수행하면 동일 상품을 파는 판매자들이 한 클러스터에 모이게 된다. 한 예로 상품 m_3 를 판매하는 판매자 s_1, s_3, s_4, s_5 는 하나의 클러스터를 이루고 있다. 상품 m_3 를 기준으로 묶인 판매자들의 집합을 $I_3 = \{s_1, s_3, s_4, s_5\}$ 로 정의한다.

The process of isolating the seller score from the buyer's rating is as follows. Firstly, we cluster sellers based on the item. Figure 2 shows an example of *initial seller clustering based on the item*. In Figure 2, there are five sellers ($s_1 - s_5$) and five items ($m_1 - m_5$). The sellers who sell the same item belong to the same cluster. For example, the sellers with m_3 , who are s_1, s_3, s_4 , and s_5 , belong to the same cluster. The set of sellers who sell m_3 is represented as $I_3 = \{s_1, s_3, s_4, s_5\}$.

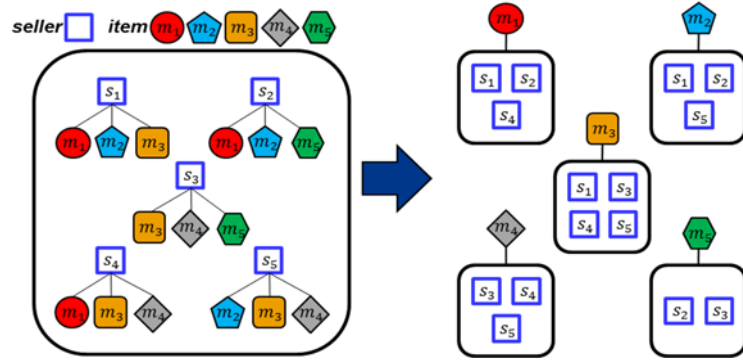


그림 2. 초기 판매자 클러스터링의 예

Fig 2: An example of initial seller clustering based on the item.

동일 상품 기준으로 묶인 판매자 클러스터들의 집합을 C^s 로 정의한다. C^{s_i} 는 e-마켓플레이스내의 i 번째 판매자 s_i 를 포함하는 모든 클러스터들의 집합, C_k^s 는 판매자 클러스터 집합의 k 번째 클러스터, $C_k^{s_i}$ 는 판매자 s_i 를 포함하는 클러스터들 중 k 번째 클러스터를 의미한다. 예를 들어 그림 2에서 $C^s = \{I_1, I_2, I_3, I_4, I_5\}$ 이며, $C^{s_2} = \{I_1, I_2, I_5\}$, $C_3^s = \{I_3\}$, $C_3^{s_2} = \{I_5\}$ 이다.

Let C^s denote the set of all seller clusters. C^{s_i} denotes the set of seller clusters, each of which includes seller s_i denoting the i -th seller in e-marketplace. For notational convenience, we use C_k^s to denote the k -th cluster in C^s and $C_k^{s_i}$ to denote the k -th cluster in C^{s_i} . For example, in Figure 2, $C^s = \{I_1, I_2, I_3, I_4, I_5\}$, $C^{s_2} = \{I_1, I_2, I_5\}$, $C_3^s = \{I_3\}$, and $C_3^{s_2} = \{I_5\}$.

둘째, 각 판매자 클러스터내의 모든 판매자들에 대해 판매자간의 상대적 점수 (판매자 점수)를 구한다. k 번째 클러스터 C_k^s 내의 판매자 s_i 의 *시작 판매자 평판*을 e_i 로 정의한다. 시작 판

매자 평판은 k 번째 클러스터를 구성할 때 기준이 되는 상품을 판매하는 판매자 s_i 에 대한 소비자들의 평점의 평균이다⁴. k 번째 클러스터 내의 모든 판매자들의 시작 판매자 평판 집합은 \mathbf{E}_k 로 정의한다. k 번째 클러스터 내의 판매자 s_i 의 판매자 점수 e_i' 는 다음과 같이 계산 할 수 있다.

Secondly, after initial seller clustering, we compute the relative scores among sellers in each cluster (i.e., seller scores). Let e_i denote the *initial seller reputation* of seller s_i in k -th cluster \mathbf{C}_k^s . In the beginning, the initial seller reputation is the average of buyers' ratings on s_i in \mathbf{C}_k^s ⁵. Let \mathbf{E}_k denote the set of initial seller reputations in \mathbf{C}_k^s . The relative score for s_i in \mathbf{C}_k^s (e_i') is computed as follows:

$$e_i' = e_i - \text{Avg}(\mathbf{E}_k \setminus e_i) \quad (\text{식 1})$$

식 1의 e_i' 는 k 번째 클러스터의 판매자 s_i 의 시작 판매자 평판 e_i 와 판매자 s_i 의 시작 판매자 평판을 제외하고 측정한 시작 판매자 평판들의 평균인 $\text{Avg}(\mathbf{E}_k \setminus e_i)$ 와의 차이이다. 각 판매자 클러스터의 모든 판매자는 판매자간의 상대적 점수인 판매자 점수 e_i' 를 갖게 된다.

e_i' in equation 1 is the seller score of s_i in \mathbf{C}_k^s and is defined as the initial seller reputation of s_i in \mathbf{C}_k^s (e_i) minus the average of the initial seller reputations of the sellers in \mathbf{C}_k^s except seller s_i ($\text{Avg}(\mathbf{E}_k \setminus e_i)$). Each seller is assigned with seller score, e_i' , within the cluster he belongs to. Note that the seller who belongs to multiple clusters would have multiple seller scores.

셋째, 최종적으로 e-마켓플레이스내의 k 번째 판매자 s_i 의 평판 e_i'' 를 구한다. e_i'' 는 자신이 포함된 모든 판매자 클러스터에서 부여 받은 판매자 점수들의 평균 값으로 한다. 아래 식에서 e_i' 는 s_i 를 포함하는 모든 클러스터 중 k 번째 클러스터에 포함된 판매자 s_i 의 판매자 점수다.

Finally, the seller reputation of seller s_i in an e-marketplace, e_i'' , is the average of the seller scores from all the clusters the seller belongs to. e_i'' is computed as in Equation 2. e_i'' is normalized to e_i^* within the range of $[0, 1]$ by min-max normalization in Equation 3.

⁴ 시작 판매자 평판을 구할 때 BSD를 적용 시킬 수 있다. BSD를 적용할 경우, e_i 는 k 번째 클러스터를 구성할 때 기준이 되는 상품(들)을 판매하는 판매자 s_i 에 대한 소비자들의 평점과 그 상품이 속하는 상품군 g 에서의 평점 신뢰도 t_r 의 가중치의 합으로 결정된다.

⁵ When computing the initial seller's reputations, BSD can be applied. When using BSD, e_i is the average of ratings weighted by the trustworthiness of the ratings given to s_i . This is explained in Section 3.3 in more detail.

$$e_i'' = \frac{1}{|C^{s_i}|} \sum_{c_k^{s_i} \in C^{s_i}} e_i' \quad (\text{식 2})$$

e-마켓플레이스내의 판매자 s_i 의 판매자 점수 e_i'' 는 식 3의 최대-최소 정규화 (min-max normalization)를 통해 $[0, 1]$ 범위내의 값인 e_i^* 로 변환하여 사용된다.

$$e_i^* = \text{Normalized}(e_i'') = \frac{e_i'' - \text{Min}_{s_i}(e_i'')}{\text{Max}_{s_i}(e_i'') - \text{Min}_{s_i}(e_i'')} \quad (\text{식 3})$$

위와 같이, 동일 상품 판매자의 상대적인 점수 차이를 이용하여 판매자의 평판을 구하는 방법을 *판매자 평점 분리* (*Seller Rating Separation: SRS*)라고 부른다. 특히, 초기 판매자 클러스터링을 수행하는 최초의 SRS를 *Initial SRS*이라 명칭하고, 이후의 SRS와 구분한다⁶.

The computation process, described above, that calculates the seller's reputation using the relative scores among the sellers with same item is called *Seller Rating Separation (SRS)*. Especially, we call the SRS which performs as an initial seller clustering *Initial-SRS* and distinguish it from subsequent SRSs⁷.

IRS는 다음과 같이 진행된다. 첫째, SRS를 통해 계산된 판매자 평판을 기반으로 상품 클러스터링을 수행한다. 그림 3은 판매자 평판기반 상품 클러스터링에 대한 예이다. 그림 3의 예제에도 총 다섯 명의 판매자 ($s_1 \sim s_5$)와 다섯 개의 상품 ($m_1 \sim m_5$)이 존재한다. 상품 클러스터링을 수행하면 판매자 평판이 유사한 판매자들의 판매 상품들이 하나의 클러스터를 형성한다. 본 논문에서는 판매자간 판매자 점수 차이가 ε 이하 일 때 유사하다고 정의한다. 그림 3의 예시에서 ε 는 0.05이며, 평판의 차이가 0.05 이내인 s_1 과 s_3 의 상품들이 하나의 클러스터를 형성한다. 그 밖의 판매자들인 s_2, s_4, s_5 는 유사 평판을 가진 판매자들이 없기 때문에 각각 자신의 판매 상품들만으로 클러스터를 구성한다. 판매자 s_1 과 s_3 를 기준으로 묶인 상품들의 집합을 $L_1 \cdot L_3 = \{m_1, m_2, m_3, m_4, m_5\}$ 로 정의한다.

The process of IRS is as follows. Firstly, we cluster items based on the sellers' reputations attained by Initial-SRS (or SRS in subsequent iterations). The items with similar sellers' reputations are clustered together. The term 'similar' means that the difference between two sellers' reputations is less than ε . Figure 3 shows an example of item clustering. In Figure 3, there are

⁶ 최초의 판매자 클러스터링의 경우에 한에서만 동일 상품 기준으로 판매자들을 클러스터링하며 이후의 모든 판매자 클러스터링들은 계산된 상품 평판을 기반으로 수행된다. Initial SRS를 수행하면 상품 기준으로 판매자가 묶이지만, 평점 분리 (RS)를 반복 수행하면서 SRS 할 경우 동일 상품 기준이 아닌 상품 평판이 유사한 상품들이 기준이 되어 클러스터가 형성되기도 한다.

⁷ Only when doing an initial seller clustering, we cluster sellers based on the same item. After that, subsequent seller clusterings are performed based on the item reputations newly attained by IRS.

five sellers ($s_1 - s_5$) and five items ($m_1 - m_5$), same as in Figure 2. After item clustering, the items sold by sellers whose reputations are similar belong to the same cluster. In Figure 3, where ε is set as 0.05, the items sold by s_1 or s_3 belong to the same cluster because the difference between s_1 's reputation (0.2) and s_3 's reputation (0.23) is less than 0.05. We denote the item cluster that includes the items sold by either s_1 or s_3 as $\mathbf{L}_1\mathbf{L}_3 = \{m_1, m_2, m_3, m_4, m_5\}$. If a seller's reputation is not similar to other sellers', the items sold by that seller belong to an individual cluster. In Figure 3, for example, items sold by s_5 (m_2, m_3 , and m_4) belong to a single cluster based on s_5 .

유사 평판 판매자 기준으로 묶인 상품 클러스터들의 집합을 \mathbf{C}^m 로 정의한다. \mathbf{C}^{mj} 는 e-마켓플레이스내의 j 번째 상품 m_j 를 포함하는 모든 클러스터들의 집합, \mathbf{C}_k^m 는 상품 클러스터 집합의 k 번째 클러스터, \mathbf{C}_k^{mj} 는 상품 m_j 를 포함하는 클러스터들 중 k 번째 클러스터를 의미한다. 예를 들어 그림 3에서 $\mathbf{C}^m = \{\mathbf{L}_1 \cdot \mathbf{L}_3, \mathbf{L}_2, \mathbf{L}_4, \mathbf{L}_5\}$ 이며, $\mathbf{C}^{m_1} = \{\mathbf{L}_1 \cdot \mathbf{L}_3, \mathbf{L}_2, \mathbf{L}_4\}$, $\mathbf{C}_1^m = \{\mathbf{L}_1 \cdot \mathbf{L}_3\}$, $\mathbf{C}_3^{m_1} = \{\mathbf{L}_4\}$ 이다.

Let denote \mathbf{C}^m the set of all item clusters. \mathbf{C}^{mj} the set of item clusters, each of which includes j -th item m_j in e-marketplace. For notational ease, we use \mathbf{C}_k^m to denote the k -th cluster in \mathbf{C}^m and \mathbf{C}_k^{mj} to denote the k -th cluster in \mathbf{C}^{mj} . For example, in Figure 3, $\mathbf{C}^m = \{\mathbf{L}_1\mathbf{L}_3, \mathbf{L}_2, \mathbf{L}_4, \mathbf{L}_5\}$, $\mathbf{C}^{m_1} = \{\mathbf{L}_1\mathbf{L}_3, \mathbf{L}_2, \mathbf{L}_4\}$, $\mathbf{C}_1^m = \{\mathbf{L}_1\mathbf{L}_3\}$, and $\mathbf{C}_3^{m_1} = \{\mathbf{L}_4\}$.

둘째, k 번째 클러스터 \mathbf{C}_k^m 내의 상품 m_j 의 시작 상품 평판을 z_j 로 정의하며, k 번째 클러스터 내의 모든 상품들의 시작 상품 평판 집합은 \mathbf{Z}_k 로 정의한다. 시작 상품 평판 z_j 는 k 번째 클러스터를 구성할 때 기준이 되는 판매자(들⁸)가 판매하는 상품 m_j 에 대한 소비자들의 평점의 평균이다. 예를 들어 그림 3에서 $\mathbf{L}_1 \cdot \mathbf{L}_3$ 에 속하는 상품 m_1 의 z_1 은 m_1 을 파는 판매자 s_1 과 s_3 이 받은 모든 평점의 평균을 의미한다. k 번째 클러스터 내의 상품 m_j 의 상품 점수 z_j' 를 다음과 같이 계산한다.

Secondly, we compute the relative scores among items in each cluster (i.e., item scores). Let z_j denote the *initial item reputation* of item m_j in k -th cluster \mathbf{C}_k^m and \mathbf{Z}_k denote the set of initial item reputations in \mathbf{C}_k^m . In the beginning, the initial item reputation is the average of buyers' ratings on m_j in \mathbf{C}_k^m . For example, in Figure 3, z_1 in item cluster $\mathbf{L}_1\mathbf{L}_3$ is the average of buyers' ratings on m_1 sold by either s_1 or s_3 . The relative item score for m_j in \mathbf{C}_k^m (z_j') is computed as

⁸ 상품 클러스터링의 경우 판매자 간의 점수 차이가 ε 는 이하인 판매자는 유사하다고 가정한다. 평점 분리 (RS)를 반복 수행하면서 상품 클러스터링을 할 경우, 판매자 점수가 유사한 판매자들이 파는 상품들이 묶이기도 유사 판매자 점수를 갖는 판매자가 없는 판매자는 자신의 상품들만 묶이기도 한다.

follows:

$$z'_j = z_j - \text{Avg}(\mathbf{Z}_k \setminus z_j) \quad (\text{식 4})$$

식 4의 z'_j 는 k 번째 클러스터의 상품 m_j 의 시작 상품 평판과 상품 m_j 의 시작 상품 평판을 제외하고 측정한 시작 상품 평판들의 평균인 $\text{Avg}(\mathbf{Z}_k \setminus z_j)$ 와의 차이이다. 각 상품 클러스터의 모든 상품은 상품간의 상대적 점수인 상품 점수 z'_j 를 갖게 된다.

z'_j in equation 4 is the item score of m_j in \mathbf{C}_k^m and is defined as the initial reputation of m_j in \mathbf{C}_k^m (z_j) minus the average of the initial item reputations of the items in \mathbf{C}_k^m except z_j ($\text{Avg}(\mathbf{Z}_k \setminus z_j)$). Each item is assigned with item score, z'_j , within the cluster that each item belongs to. Note that an item which belongs to multiple clusters would have multiple item scores.

셋째, 최종적으로 e-마켓플레이스내의 j 번째 상품 m_j 의 평판 z''_j 를 구한다. z''_j 는 자신이 포함된 모든 상품 클러스터에서 부여 받은 상품 점수들의 평균 값으로 한다. 아래 식에서 z'_j 는 m_j 를 포함하는 모든 클러스터 중 k 번째 클러스터에 포함된 상품 m_j 의 점수다. e-마켓플레이스내의 상품 m_j 의 판매자 점수 z''_j 는 식 3과 마찬가지로 최대-최소 정규화를 통해 $[0, 1]$ 범위내의 값인 z^*_j 로 변환하여 사용된다.

Finally, the item reputation of item m_j in an e-marketplace denoting z''_j is the average of the item scores from all the clusters that the item belongs to. z'_j is computed as in equation 5. z''_j is normalized to z^*_j within the range of $[0, 1]$ by the min-max normalization as same as Equation 3.

$$z''_j = \frac{1}{|\mathbf{C}^{m_j}|} \sum_{\mathbf{C}_k^{m_j} \in \mathbf{C}^{m_j}} z'_j \quad (\text{식 5})$$

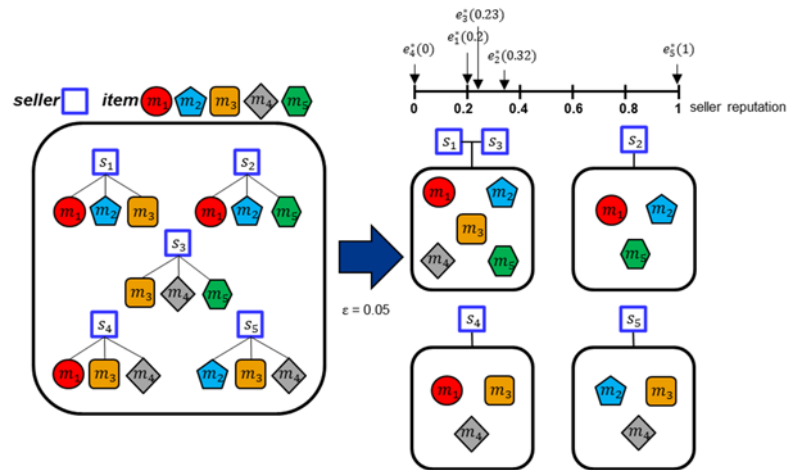


그림 3. 상품 클러스터링의 예

Fig 3: An example of item clustering based on sellers' reputations.

그림 4는 IRS와 동일한 메커니즘으로 동작하는 일반적인 SRS의 예이다. 이전 IRS를 통해 도출된 각 상품의 평판을 m_1 은 1, m_2 는 0.44, m_3 는 0.39, m_4 는 0, m_5 는 0.14라고 하자. ε 는 마찬가지로 0.05이며, 상품 평판의 차이가 0.05이내인 m_2 과 m_3 를 판매하는 판매자들이 하나의 클러스터를 형성한다.

Figure 4 is an example of SRS. The process of SRS is similar to that of IRS. Assume that, after IRS, item reputations of m_1, m_2, m_3, m_4 , and m_5 are computed as 1, 0.44, 0.39, 0, and 0.14, respectively. When ε is assumed to be set as 0.05, sellers selling both m_2 and m_3 belong to the same cluster.

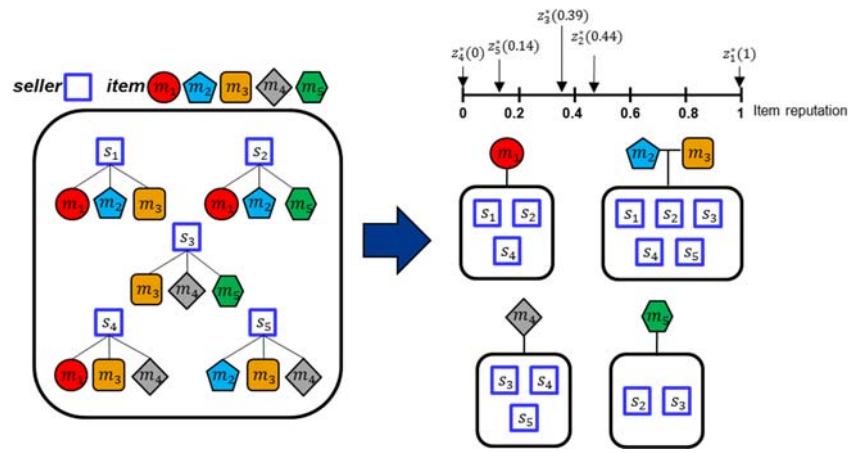


그림 4. 판매자 클러스터링의 예
Fig 4: An example of seller clustering based on the item.

3.2 Better Seller Discovery (BSD)

평점의 신뢰도는 **활동성**, **다양성**, **대중성**이라는 세가지 핵심 요소를 사용하여 계산한다. 활동성은 e-마켓플레이스 내에 소비자의 활동량을 나타내는 값이다. 다양성은 소비자가 얼마나 다양한 판매자와 거래를 하였는지를 나타내는 값이다. 대중성은 특정 판매자에 대한 어떤 소비자의 평가가 다른 소비자들 (대중)의 평가에 얼마나 근접하는지를 나타내는 값이다. 저자들은 e-마켓플레이스내에서 활동량이 많고, 다양한 판매자와 거래를 수행하여, 다른 소비자와 일관된 평가를 수행한 대중성이 높은 소비자의 평점은 신뢰할 수 있다고 간주하고, 판매자 평판을 보정할 때 가중치를 부여한다. 활동성, 다양성, 대중성에 대한 자세한 내용은 다음과 같다.

The trustworthiness of a rating is computed based on three factors: activity, diversity, and universality. Activity measures the amount of buyer actions. Diversity captures the extent to which a buyer has transacted with various sellers. Universality quantifies how conforming the rating of a buyer is to the majority opinion. We consider the rating more trustable when the buyer has rated many sellers, the buyer has transacted with more diverse sellers, and the rating is

conforming to majority opinions. Adjusting the original ratings by assigning higher weights to the ratings of trustable buyers, we re-compute the sellers' reputations. Activity, diversity, and universality are defined more formally in the following.

활동성 e-마켓플레이스에는 다양한 평가를 통해 자신의 경험을 공유하려는 소비자들이 있는 반면에 상품을 구매했음에도 불구하고 평가를 일절 하지 않는 소비자도 존재한다. 소비자 활동성은 e-마켓플레이스 내에 소비자의 활동량을 나타내는 값으로 소비자의 평점 빈도를 이용하여 수치화한다.

Activity We measure the activity of a buyer in an e-marketplace based on the number of evaluations the buyer has performed (i.e., the number of transaction he rates).

다양성 소비자 다양성은 소비자가 얼마나 다양한 판매자와 거래 하였는지를 나타내는 값이다. 상대적으로 다양한 판매자와 거래를 수행한 소비자는 그렇지 않은 소비자에 비해 다수의 판매자를 비교하여 평가할 수 있기 때문에 평가에 상대성이 충분히 반영되어 있다. 극소수의 판매자와만 협소한 거래를 진행한 소비자의 경우, 비교 평가의 대상이 되는 판매자가 부족하여 소비자의 평가에 상대성이 결여되어 있음은 물론 판매자의 행동에 상관없이 우호적인 평가 일변도 일 수 있다. 즉, 다양성이 결여된 소비자의 평점은 신뢰도가 떨어진다고 볼 수 있다.

Diversity The diversity score is higher when a buyer has transacted with diverse sellers. The diversity score, when used in measuring the trustworthiness of a rating, captures two facts. First, a buyer who has transacted with more sellers would be more objective and trustable, because his evaluations would have been attained by comparing many sellers' behaviors. Second, a buyer who has transacted with only a small number of sellers is likely to be less objective and untrustworthy, because his evaluations would have not been. In case of bribed buyers, they would have mainly transacted with the bribing sellers in order to give unfairly positive ratings to them. Therefore, we assign a low diversity score to buyers who have transacted with a small number of sellers, while assigning a high diversity score to buyers who have transacted with diverse sellers.

e-마켓플레이스에는 자신의 평판을 인위적으로 높이기 위해 소비자들을 매수하는 판매자가 존재한다. 매수 당한 소비자는 공모 판매자에게 편중된 다수의 거래를 진행 한 후 높은 평점을 매기거나 경쟁 판매자들과 다수의 거래를 진행한 후 낮은 평점을 매겨 거짓된 정보를 양산한다. 즉, 매수된 소비자는 소수의 판매자와만 거래를 하는 다양성이 결여된 소비자로 볼 수 있다. 따라서, 저자들은 소비자가 특정 판매자와 편중된 거래를 하였을 경우 낮은 다양성을 부여하고, 상대적으로 다수의 판매자와 다양한 거래를 하였을 경우 높은 다양성을 부여한다.

대중성 소비자의 대중성은 특정 판매자에 대한 어떤 소비자의 평가가 다른 소비자들 (대중)의

평가와 얼마나 유사한지를 나타내는 값이다. '평점'은 거래에 대한 소비자 개인의 만족도를 수치화한 주관적인 값이지만 '평판'은 다양한 소비자 평점의 집합체로서 거래에 대한 대중들의 전반적인 만족도를 나타내는 객관적인 값이라고 할 수 있다. 따라서 판매자에 대한 (보다 객관적인) 대중의 평가 (평판)와 (보다 주관적인) 소비자의 평가 (평점)간의 차이 값을 이용하여 대중성을 측정한다. 소비자의 대중성은 소비자의 각 평점이 평판과 유사할수록 증가하고, 평점과 평판의 차이가 클수록 감소한다.

Universality A buyer is considered having universality if his evaluation (i.e., rating) is closer to the majority opinion (i.e., reputation). Universality is measured based on the deviation of the buyer's ratings from the reputations of the sellers. The more similar a buyer's rating is to the reputation, the higher the universality of the buyer; the more dissimilar, the lower.

평점의 신뢰도는 평점을 작성한 구매자의 활동성, 다양성, 대중성을 계량화한 값들의 연산으로 계산된다. 단, 구매자는 각기 다른 상품 선호도를 가지고 있기 때문에 상품 군⁹별로 평점 이력의 패턴이 다를 수 있다. 예를 들어, 어떤 소비자는 전자제품 군에 특별한 관심을 갖고 있어 다수의 전자제품 거래를 성사시킨 반면에 가구 상품 군이나 옷 상품 군에는 큰 관심이 없어 한 건의 거래도 진행하지 않을 수 있다. 이 경우 해당 소비자는 전자제품 군에는 정통하다 할 수 있기 때문에 해당 상품 군에 대한 소비자의 평점들을 신뢰 할 수 있는 반면, 그 밖의 상품 군에 대한 소비자의 평가는 상대적으로 신뢰하기 어렵다. 따라서 각 상품 군 별로 구매자의 평점 이력을 분석하여 구매자의 활동성, 다양성, 전문성을 다르게 측정한다. 신뢰도를 구하는 방법과 구해진 신뢰도를 이용하여 평판을 보정하는 방법을 이후에 자세히 설명한다. 표 1은 본 절에서 사용되는 기호들이다.

Note that because each buyer has different interests, he may have a different purchase/rating history according to an item group¹⁰. For example, a buyer who has been actively purchasing and evaluating for new books may not have purchased anything from cosmetic categories. In this case, the buyer should be assigned a high score for activity, diversity, and universality in the book category but not in the cosmetic category. Therefore, it is more reasonable to measure activity, diversity, and universality within each item group. We describe below how to compute

⁹ 본 논문에서 상품 군은 ebay에서 제공하는 상품 카테고리 (<http://www.ebay.com/sch/allcategories/all-categories>)중 최하위 카테고리 (예를 들면, 전자제품, 가구, 옷 등등)를 의미한다.

¹⁰ The item group refers to the lowest category of ebay.com, such as smartphone, furniture, camera, and so on (<http://www.ebay.com/sch/allcategories/all-categories>). .

the trustworthiness of ratings and how to re-compute seller's reputations using trustworthiness of ratings. The notations used in this section are presented in Table 1.

표 1. 기호들

Table 1. Notations

<i>Name</i>	<i>Description</i>
U	set of buyers
S	set of sellers
M	set of items
R	set of ratings
G	set of item groups
u	a buyer, $u \in \mathbf{U}$
s	a seller, $s \in \mathbf{S}$
\bar{s}	reputation of seller
m	an item, $m \in \mathbf{M}$
r	a rating, $r \in \mathbf{R}$
r_m	a rating to item m
g	an item group, $\mathbf{g} \subset \mathbf{G}$, $m \in \mathbf{g}$
\mathbf{U}^g	set of buyers in g
\mathbf{R}^g	set of ratings in g
\mathbf{R}_u^g	set of ratings by u in g , $\mathbf{R}_u^g \subset \mathbf{R}^g$
\mathbf{R}_s	set of ratings given to s
\mathbf{R}_s^g	set of ratings given to s in g , $\mathbf{R}_s^g \subset \mathbf{R}^g$
\mathbf{S}_u^g	set of sellers rated by u in g

3.2.1 활동성 Activity

소비자 활동성은 상품 군 \mathbf{g} 에서의 소비자의 평점 빈도를 이용하여 측정한다. 특정 상품 군에서 어떤 소비자의 평점 빈도가 유달리 높다면 해당 상품 군에 대한 관심도가 높은 활동적인 소비자로 간주하고 해당 소비자의 활동성을 높게 측정한다. 상품 군 \mathbf{g} 에서의 소비자의 활동성 a_u 는 다음 식(6)과 같이 정의된다.

The activity of buyer u , denoted by a_u , is quantified by the frequency of his ratings in an item group \mathbf{g} . If a buyer has given many ratings in an item group, we regard him as an active buyer in that item group and assign a high activity score to him. We define a_u in \mathbf{g} as follows:

$$a_u = |\mathbf{R}_u^g| - \frac{|\mathbf{R}^g|}{|\mathbf{U}^g|} \quad (\text{식 6})$$

$$a_u^* = \text{Normalized}(a_u) = \frac{a_u - \text{Min}(a_u)}{\text{Max}(a_u) - \text{Min}(a_u)} \quad (\text{식 7})$$

소비자 u 의 활동성 (activity) a_u 은 어떤 상품 군 g 에서 u 가 다른 소비자들에 비해 상대적으로 얼마나 더 많은 평가를 수행하였는지를 수치화 한 값이다. a_u 는 g 에 속한 상품들에 대해 u 가 매긴 평점 개수 $|\mathbf{R}_u^g|$ 와 g 에 1회 이상 평점을 매긴 모든 소비자들의 평균 평점 개수 즉, g 에 대한 모든 평점 개수 $|\mathbf{R}^g|$ 를 g 에 1회 이상 평점을 매긴 모든 소비자의 수 $|\mathbf{U}^g|$ 로 나눈 값의 차이로 정한다. 상품 군 g 에서의 소비자 u 의 활동성 a_u 은 식 7의 최대-최소 정규화 (min-max normalization)를 통해 $[0, 1]$ 범위내의 값인 a_u^* 로 변환하여 사용된다. 이때 $\text{Max}(a_u)$ 는 상품 군 g 에서 가장 높은 활동성을 $\text{Min}(a_u)$ 는 상품 군 g 에서 가장 낮은 활동성을 의미한다.

A buyer u who gives more ratings than normal buyers in item group g should be considered as an active buyer in g . a_u is the difference between the frequency of ratings by u in g , denoted by $|\mathbf{R}_u^g|$, and the average of rating frequencies of all buyers in g , which is $|\mathbf{R}^g|$ divided by $|\mathbf{U}^g|$. In order to use a_u in calculating the trustworthiness of ratings, we normalize a_u to a_u^* in the range of $[0, 1]$ by applying min-max normalization, similar to equation (7). $\text{Max}(a_u)$ indicates the highest activity in g , and $\text{Min}(a_u)$ the lowest activity in g .

3.2.2 다양성 Diversity

소비자 u 의 다양성 (diversity) d_u 은 상품 군 g 에서의 판매자에 대한 구매자 평점 분포를 이용하여 측정한다. d_u 는 상품 군 g 에서 u 가 특정 판매자와 집중적으로 거래했을 때 그 값이 낮아지는 반면 다양한 판매자와 거래 했을 때 그 값이 높아진다. 상품 군 g 에서의 소비자의 다양성 d_u 는 다음과 같이 정의된다.

The diversity of buyer u , denoted by d_u , is measured by analyzing the distribution of his ratings in g . If the buyer has transacted with only a small number of sellers in g , his diversity score is lower, while with diverse sellers, it is higher. We define d_u in g as follows:

$$d_u = \frac{|\mathbf{S}_u^g|}{|\mathbf{R}_u^g|} \quad (\text{식 8})$$

d_u 는 g 에서 u 가 거래한 (평점을 매긴) 모든 판매자의 수 $|\mathbf{S}_u^g|$ 를 g 에 u 가 매긴 모든 평점 개수인 $|\mathbf{R}_u^g|$ 로 나눈 값이다. 만약 g 에서 u 가 거래한 판매자의 수와 u 가 매긴 모든 평점 개수가 동일하다면 d_u 는 1로 측정된다. 만약 g 에서 u 가 거래를 시도한 판매자가 단 1명이라면 d_u 값은 u 의 평점 개수가 증가함에 따라 그 값에 0에 가까워지게 된다. 평점 신뢰도 계산을 위해

상품 군 \mathbf{g} 에서의 소비자 다양성 d_u 도 식 7과 같이 최대-최소 정규화를 이용해서 $[0, 1]$ 범위내의 값인 d_u^* 로 변환한다. 이때 $Max(d_u)$ 는 상품 군 \mathbf{g} 에서 가장 높은 다양성을 $Min(d_u)$ 는 상품 군 \mathbf{g} 에서 가장 낮은 다양성을 의미한다.

d_u is the number of the sellers rated by buyer u in \mathbf{g} ($|S_u^g|$) divided by the number of ratings by u in \mathbf{g} ($|R_u^g|$). If the number of ratings by u is same as the number of the sellers rated by u , d_u is 1. If u has done most of his transactions with only a smaller number of sellers, d_u will be close to 0. We normalize d_u to d_u^* in the range of $[0, 1]$ by applying min-max normalization. $Max(a_u)$ indicates the highest diversity in \mathbf{g} , and $Min(a_u)$ the lowest diversity in \mathbf{g} .

3.2.3 대중성 Universality

대중성은 특정 판매자에 대한 어떤 소비자의 평가가 다른 소비자들 (대중)의 평가에 근접하는지를 나타내는 값이다. ~~소비자 u 의 대중성 (universality) p_u 은 상품 군 \mathbf{g} 내의 어떤 특정 판매자 한명이 판매하는 하나의 상품 m 거래에 대한 u 의 평점 r_m 과 (그 특정 판매자 한명이 파는 하나의) 상품 m 에 대한 모든 소비자들의 평균 평점 \bar{r}_m 의 근접성 측정을 기반으로 도출된다.~~ 즉, 대중성은 소비자의 r_m 이 \bar{r}_m 에 근접할수록 높아진다. r_m 의 근접성은 \bar{r}_m 에 대한 r_m 의 표준점수의 절대값으로 측정되며 이때 추가적으로 m 에 매겨진 모든 평점들에 대한 표준편차인 s_m 이 사용된다. r_m 의 근접성 r'_m 는 다음과 같이 정의된다.

If a buyer's rating conforms to the majority of other buyers' ratings, he should be considered having universality. University is measured by the average of proximities of the buyer's ratings. Proximity is computed based on the deviation of buyer u 's rating on item m in \mathbf{g} , denoted by r_m , from the average of the ratings on m in \mathbf{g} , denoted by \bar{r}_m . The more similar are r_m and \bar{r}_m , the higher is the universality of u , and the more dissimilar they are, the lower the universality of u . The proximity of r_m , denoted by p_m , is calculated based on \bar{r}_m and the standard deviation, denoted by s_m , as follows:

$$p_m = \left| \frac{r_m - \bar{r}_m}{s_m} \right| \quad (\text{식 9})$$

식 9을 통해 계산된 r'_m 은 그 값이 0에 가까울수록 r_m 을 매긴 소비자 u 가 다른 소비자들과 일관된 평점을 매겼음을 의미한다. 소비자의 대중성 p_u 은 소비자가 \mathbf{g} 에 매긴 모든 평점 근접성의 평균 값으로 계산되며, 그 값이 0에 가까울수록 소비자가 다른 소비자들 (대중)과 유사하게 평점을 매긴 대중성이 높은 소비자임을 의미한다. p_u 는 다음과 같이 정의된다.

If p_m is closer to 0, it indicates that u 's opinion for m is consistent with those of other buyers. University, denoted by v_u , is calculated as the average of all the proximities (p_m s) of the ratings

by u in \mathbf{g} . If v_u is closer to 0, it indicates that he has a high universality conforming to the majority of other buyers' opinions. We define v_u as follows:

$$v_u = \frac{1}{|\mathbf{R}_u^g|} \sum_{r_m \in \mathbf{R}_u^g} p_m \quad (\text{식 } 10)$$

소비자의 대중성 v_u 을 평점 신뢰도 계산에 이용하기 위해서는 정규화 과정이 필요하다. 평점 신뢰도는 그 값이 크면 클수록 신뢰할 수 있는 평점임을 의미하기 때문에 v_u 가 0에 가까울수록 실제로는 보다 큰 양의 값으로 변환되어야 실제 평점 신뢰도 계산에 쉽게 이용할 수 있다. 활동성, 다양성과 마찬가지로 평점 신뢰도 계산을 위해 최대-최소 정규화를 이용해서 $[0, 1]$ 범위내의 값으로 변환하되, 변환한 값과 1과의 차이를 최종 변환 값 v_u 로 사용한다. 이때 $\text{Max}(p_u)$ 는 상품 군 \mathbf{g} 에서 가장 높은 대중성을 $\text{Min}(p_u)$ 는 상품 군 \mathbf{g} 에서 가장 낮은 대중성을 의미한다.

In order to use v_u in calculating the trustworthiness of ratings, we need to normalize v_u . Because the high scores of the trustworthiness of ratings means the more trustable ratings, the universality score used in calculating the trustworthiness of ratings should be close to 1 when v_u moves closer to 0. We normalize v_u to v_u^* in the range of $[0, 1]$ by applying min-max normalization and also computing the difference one minus the normalized value like equation 11. $\text{Max}(v_u)$ indicates the highest universality in \mathbf{g} and $\text{Min}(v_u)$ the lowest universality in \mathbf{g} .

$$v_u^* = \text{Normalized}(v_u) = 1 - \frac{v_u - \text{Min}(v_u)}{\text{Max}(v_u) - \text{Min}(v_u)} \quad (\text{식 } 11)$$

3.2.4 신뢰도 계산 Trustworthiness Computation

BSD는 상품 군별로 소비자의 활동성, 다양성, 대중성을 계산하여 상품 군내에 모든 평점의 신뢰도를 측정하고, 측정된 평점 신뢰도를 기반으로 판매자의 평판을 보정한다. 상품 군 \mathbf{g} 에서의 평점 신뢰도는 다음과 같이 계산된다.

BSD measures the trustworthiness of all ratings in item group \mathbf{g} using the activity, diversity, and universality scores of a buyer, and re-computes seller reputations. The trustworthiness of a rating in item group \mathbf{g} , denoted by t_r , is computed as follows:

$$t_r = a_u^* \times d_u^* \times v_u^*, \quad r \in \mathbf{R}_u^g, u \in \mathbf{U}^g \quad (\text{식 } 12)$$

판매자의 평판을 계산할 때 판매자의 평판은 기존의 산술평균이 아닌 평점의 신뢰도를 가중치로 이용한 가중치 평균으로 보정된다. 다시 말해, 판매자 s 의 평판 \bar{s} 는 s 에게 평가를 매긴 모든 구매자 \mathbf{U}^s 의 평점과 각 구매자의 평점이 상품 군 \mathbf{g} 에서 갖는 신뢰도 t_r 를 가중치로 이용하는 가중치 평균으로 보정되며 (이 말도 필요합니다. 왜 중요한 말들을 ambiguous하고 generic한 표현으로 바꾸나요? 이런 식으로 하면 독자들은 영어를 읽어도 무슨 뜻인지 정확히

파악 할 수 없습니다.) 식은 다음과 같다.

Instead of using the average (i.e., to assign the same weight to each rating), we re-compute the reputation of a seller s as the average of ratings weighted by the trustworthiness (t_r) of ratings given to s . To be more concrete, because the reputation of a seller is computed separately according to item group \mathbf{g} , the seller (s) reputation in e-marketplace, denoted by \bar{s} , is computed as the average of ratings in \mathbf{U}^s adjusted based on all the t_r s of ratings to the seller, which are already differently computed according to \mathbf{g} which each rating belong to as follows:

$$\bar{s} = \sum_{\mathbf{R}_s^g \subset \mathbf{R}_s} \sum_{r \in \mathbf{R}_s^g} \frac{r \times t_r}{t_r} \quad (\text{식 13})$$

평점 신뢰도 계산과 계산된 신뢰도 기반의 판매자 평판 보정 방법을 포함하는 알고리즘인 BSD는 다음과 같다.

The overall algorithm of BSD is presented below.

Algorithm 2 Better Seller Discovery (BSD)

Input: e-marketplace data: set of buyers \mathbf{U} , set of sellers \mathbf{S} , set of items \mathbf{M} , set of ratings \mathbf{R} , set of item groups \mathbf{G}

Output: reputations of all sellers

//TRUSTWORTHINESS COMPUTATION

for each $\mathbf{g} \subset \mathbf{G}$

for each $u \in \mathbf{U}^g$

Compute a_u^*, d_u^*, v_u^* ,

for each $r \in \mathbf{R}^g$

Compute $t_r = a_u^* \times d_u^* \times v_u^* \ (r \in \mathbf{R}_u^g)$

//REPUTATION ADJUSTMENT

for each $s \in \mathbf{S}$

Compute $\bar{s} = \sum_{\mathbf{R}_s^g \subset \mathbf{R}_s} \sum_{r \in \mathbf{R}_s^g} \frac{r \times t_r}{t_r}$

3.3 RS내의 BSD의 활용

BSD를 RS에 접목하면 보다 신뢰할 수 있는 평판을 구할 수 있다. 첫째, SRS내의 판매자 s_i 의 시작 판매자 평판인 e_i 를 구하는 단계에서 BSD를 접목하면, unfair ratings의 영향을 줄여 평판의 신뢰를 보다 향상시킬 수 있다. (신뢰를 향상시킬 수 있는게 시작 판매자 평판만인가요? 아니면 다른 단계도 마찬가지인가요? 써놓은 것은 마치 시작 판매자 평판만인 것처럼 썼습니다.) BSD를 접목할 경우 계산은 다음과 같이 진행된다. e_i 는 판매자 클러스터를 구성할 때 기준이 되는 상품 m (또는 상품들)을 판매하는 판매자 s_i 에 대한 소비자들의 평점들과 그 상품이 속하는 상품 군 \mathbf{g} 에서의 평점 신뢰도 t_r 의 가중치의 합으로 결정된다.

By combining RS and BSD, we can attain more trustable seller reputations. First, when performing the initial seller reputations (e_i s) in initial SRS, we can reduce the influence of unfair

ratings by applying BSD. BSD can be applied to computing the initial seller reputations in subsequent SRSs. When applying BSD to SRS, in the process of computing $e_i s$, e_i is determined as the average of ratings weighed by the trustworthiness of ratings to the seller (s_i), which are computed differently according to g .

둘째, IRS내에 시작 상품 평판 z_j 를 구하는 단계에 BSD를 적용할 경우 계산은 다음과 같이 진행된다. z_j 는 상품 클러스터를 구성할 때 기준이 되는 판매자 (또는 판매자들)가 판매하는 상품 m_j 에 대한 소비자들의 평점들과 그 상품이 속하는 상품 군 g 에서의 평점 신뢰도 t_r 의 가중치의 합으로 결정된다.

Second, similar to the SRS process applied by BSD, we can compute the initial item reputations ($z_j s$) in IRS by applying BSD. This IRS process can also reduce the influence of unfair ratings. In subsequent IRSs, BSD can be applied when computing the initial item reputations. In the process of computing $z_j s$, z_j is the average of ratings weighed by the trustworthiness of ratings to the item (m_j), which are computed differently according to g .

4. Experiments

~~실제 평판 시스템에서 어떤 소비자가 어떤 판매자와 공모를 하는지에 대한 정보를 알 길이 없기 때문에 UR attackers가 존재하는 e마켓플레이스 데이터를 얻는 것은 불가능하다. 따라서 기존의 평판 시스템 분야의 연구들은 가상의 UR attackers와 가상의 e 마켓플레이스 데이터를 직접 생성하여 사용한다. 본 실험에서는 실제 e 마켓 플레이스의 복잡한 상황을 반영하는 시뮬레이션 방법을 제안하고 다양한 종류의 UR attackers를 생성하여 사용한다.~~

~~The ground truth data about UR attacks in e marketplace is unavailable because there is no way of knowing which buyers gave unfair ratings to which sellers in real world. Previous researches have generated artificial data where simple (???) transactions occur and some buyers generate unfair ratings (여기에서 정확하게 무슨 의미를 전달하려고 하나요??? 이런 의미를 전하고 싶었나요?) and use them for their simulation. In comparison, our simulation model reflects more complicated situations in e marketplace (예를 들면, 어떤 복잡한 상황을 말하나요?) and generates various UR attackers. . (문단 전체의 구성이 이렇게 되어야 더 맞을 것 같습니다. (1) 예전 연구들은 간단한 상황과 간단한 unfair ratings만 모델해서 실험했다. 간단하게 모델한 이유는 그들은 rating separation을 안하기 때문 아닌가요? Unfair rating은 원래 모델이 불가능하다. 그래서 그들은 어떻게 했다. (2) 우리 시뮬레이션에서는 복잡한 상황과 UR attacks을 모델한다. (3) 복잡한 상황은 뭐다. (4) UR attack은 체계적으로(???) 한다.)~~

Previous researches have generated artificial data where transactions based on simplified buyer-seller behavior patterns occur. Not explicitly taking items into account, they have commonly

assumed that a buyer evaluates a transaction based on the capability of a seller only. In real e-marketplace situations where various items exist and a wide variety of sellers may sell different items in different item groups, however, a buyer's evaluation of a transaction is based on two aspects of the transaction: the capability of a seller and the quality of an item. Therefore, in comparison, our simulation model reflects the tripartite relationships between buyers, sellers, and items.

Also, the ground-truth data about UR-attacks in e-marketplace is unavailable because there is no way of knowing which buyers gave unfair ratings to which sellers in real world. We artificially generates comprehensive types of UR-attacker, almost all of which examined in previous researches [7][10][11][12][13][15], who have made undesirable and negative influence on the seller selection process of potential buyers.

4.1 Description of e-marketplace simulation

~~판매자와 소비자간에 단순한 행동 양식만 정의하여 시뮬레이션을 구축한 기존 연구들과 [7, 8, 9, 10] 달리, 본 실험에서는 판매자와 소비자의 특성뿐만 아니라 상품에 대한 특성도 구체적으로 정의한 시뮬레이션을 구축하여, 이 3가지 객체의 조합에 따라 다양한 행동 양식 (behavioral pattern)이 발생하는 가상의 e 마켓플레이스 데이터를 생성하였다. 거래 단위 시간은 일 (a day)로 정하여 소비자가 1일 최대 1번을 거래할 수 있다고 가정하였다.~~

~~Whereas existing simulation models are based on simple behavior patterns between sellers and buyers, our simulation model explicitly takes items into account in addition to sellers and buyers.~~

Item 상품의 특징은 상품의 품질과 상품의 카테고리에 의해 결정 된다. 상품의 품질은 [0, 1] 범위의 임의의 값을 갖도록 설정했다. 상품의 카테고리는 계층적 구조를 이루며 최하단 카테고리를 '상품 군'이라 칭한다. 본 실험에서는 카테고리 계층 구조의 깊이를 3으로 설정하고, 상단 카테고리는 3종류, 중단 카테고리는 5종류, 하단 카테고리는 6종류로 정하여 총 90개의 상품 군이 존재하도록 설정했다. 각 상품은 3 수준 (상단, 중단, 하단)의 카테고리 정보를 포함하며, 한 상품 군에서 판매하는 상품의 개수는 전체 상품 개수를 상품 군의 개수로 나눈 값의 올림수를 넘지 않도록 했다. 예를 들어, 상품의 총 개수를 1,000로 정했을 때 하나의 상품 군에 판매되는 상품의 최대 개수는 12개이다.

Item An item is modeled by two parameters: its quality and its position in hierarchical item categories. The quality of an item is set between the range of [0, 1]. Each item keeps information about its own categories. The level of hierarchical item categories is set as 3, which is typical in real e-marketplace. The top level consists of three categories, each top category consists of five

middle categories, and each middle category consists of six bottom categories. That is, total ninety item groups exist in the model. The number of items sold in each item group is set similar to each other. ~~The number of items sold in each item group is limited not to exceed the number of all the items divided by the number of all item groups.~~ (앞 문장처럼 쓰지 말고, 각 item group에서 파는 물건의 개수는 모든 item group이 비슷하도록 설정했다라고 말해야 하지 않을까요?) For example, if the number of items sold in an e-marketplace is 1,000, the number of items which can be sold in an item group is eleven or twelve.

Seller 판매자의 특징은 판매자의 능력과 판매 상품에 의해 결정된다. 판매자의 능력은 평균을 0.5, 표준 편차를 0.25로 하는 정규 분포에 따라 임의로 설정하되 최소값을 0으로 최대값을 1로 한다. 판매자의 능력이 0에 가깝다는 것은 판매자가 불친절한 답변과 늦은 배송 등을 일삼는 것을 의미하고, 판매자의 능력이 1에 가깝다는 것은 판매자가 친절한 답변은 물론 빠른 배송 등 소비자에게 최상의 서비스를 제공하는 것을 의미한다.

Seller A seller is modeled by his capability and his selling items. The value of the capability of a seller is set based on the truncated normal distribution with an expected value of 0.5 and a standard deviation of 0.25. The truncation is set at 0 for the lower bound and 1 for the upper bound. The value close to 1 means that the seller provides fast delivery and better service, while the value close to 0 means that he provides late delivery and poor service.

e-마켓플레이스에는 다양한 상품 군이 존재하고, 판매자마다 주력 및 부 주력으로 판매하는 상품들이 다르다. e-마켓플레이스의 판매자들은 1개의 주력 상품 군과 1~3개의 부 주력 상품 군에 속하는 상품을 판매한다고 가정한다. 판매자는 1개의 주력 상품 군에 속하는 임의의 상품 3~6를 판매하고, 0~3개의 부 주력 상품 군 전체에 속하는 상품 1~6개를 임의로 판매한다.

~~Various item groups exist and each seller may sell different items in different item groups.~~ (소비자들이 관심 상품 군을 가지고 있다고 모델하는 것도 같은 맥락 아닌가요...???) (이렇게 모델한 것이 실험할 때 무슨 의미를 가지나요??? Rating separation과 관련해서 의미를 가지는 것 같은데... 그리고 이것이 앞에서 계속 말하는 complicated situation 인 것 같은데... highlight 된 이런 이야기는 앞에서 써주어야 할 것 같습니다.) (보통 셀러들은 하나의 특정 카테고리의 물품을 주로 파는 경향이 있다. 그리고 이걸 모델하는 것이 왜 중요한지를 써주어야 수궁이 갑니다.)

There exist various sellers of which some sellers sell their items only belonging to one major item group while others sell several kinds of items belonging to different item groups. In order to simulate the variety of sellers and their selling items, we model a seller as follows; each seller sells three to six items belonging to one major item group and one to six items from zero to

three minor item groups.

Buyer 소비자의 특징은 관심 상품 군에 의해 결정된다. 소비자는 3~6개의 상품 군에 이견 왜 빠졌죠??? 대해 [0, 1]사이에 임의의 관심도를 가진다. 소비자는 높은 관심도를 갖는 상품 군에 대해 더 잦은 구매를 시도한다. 거래가 완료되면 소비자는 판매자의 능력과 상품의 품질을 종합하여 평점을 매긴다.

Buyer A buyer is modeled to buy items in the item groups of his interests more often. The buyer's interest in an item group is set between [0, 1] and a buyer has his interest on three to six item groups. After completing a transaction, the buyer rates the transaction.

Transaction 소비자는 자신이 관심 있어하는 상품을 파는 판매자와 높은 확률로 거래를 수행한다. 거래 시 동일 상품을 동일한 소비자가 단기간에 구매하는 것을 방지하기 위해 상품마다 재 구매 대기 기간을 설정하여 그 기간 내에 한 소비자가 동일 상품을 사지 못하도록 하였다. 재 구매 대기 시간은 3개의 최 상단 카테고리 각각에 3일, 5일, 10일로 부여하여 각 최 상단 카테고리내 최 하단의 카테고리인 상품 군에 속하는 모든 상품에게 부여한다.

Transaction A buyer is more likely to transact with the sellers who sell the items he is interested in. (그래서 시뮬레이션에서는 동일 판매자의 같은 물건을 여러 번 사는 경우가 실제 상황보다 더 자주 생길 수 있다...인 건가요???) Therefore, in our simulation, buying the same item from the same seller repeatedly can occur more often than in real world and it is not corresponding to real e-marketplace situations. To prevent the same buyer from buying the same item from the same seller again within a short period of time, we set a limit on the repurchase time of each item, called *repurchase waiting time*. We set the repurchase waiting time of each item to be either three, five, or ten days.

모든 소비자가 동일 상품에 동일한 재 구매 대기 시간을 갖는 것을 방지하기 위해, 소비자의 상품 구매 주기를 정의한다. 상품 구매 주기는 특정 상품을 구매한 소비자가 동일 상품을 재 구매를 하는데 걸리는 최소 기간을 의미한다. 상품 구매 주기는 해당 상품의 기본 재 구매 대기 시간에 [0, 3] 범위의 임의의 값을 더해서 최종 결정된다.

~~To prevent all buyers from having the same repurchase waiting time on an item~~ (이건 또 왜 나쁜가요??? 왜 이렇게 하는지를 쓰지 않으면 이해가 안됩니다.),

If the repurchase waiting time is set only, all buyers purchasing the same item have almost identical behavior of the repurchase waiting time. To give the diversity of a purchase cycle for buyers and to prevent all buyers from having the same repurchase waiting time on an item, we define *buyer's purchase cycle*, the minimum amount of time required before a buyer can buy the

same item as the one he has already purchased. The buyer's purchase cycle on an item is computed by adding $[0, 3]$ days randomly to the repurchase waiting time of the item.

시뮬레이션에서는 상품 수, 판매자 수, 소비자 수, 전체 거래 일수가 파라미터로 사용된다. 실험 결과가 파라미터의 영향을 받는지에 대한 여부를 확인하기 위해 저자들은 2개의 서로 다른 파라미터 세트를 사용하여 e-마켓플레이스 데이터를 생성하였다. 첫 번째 파라미터 세트는 상품 1,000개, 판매자 500명, 소비자 5,000명, 전체 거래 일수 300일로 구성된다. 보다 큰 규모의 e-마켓플레이스 데이터 생성을 위해 설정한 두 번째 파라미터 세트는 상품 2,000개, 판매자 1,000명, 소비자 10,000명, 전체 거래 일수 300일로 구성된다.

(앞 문장의 설명은 added value가 없습니다. 첫번째 실험환경은 덜 활발한 것을 말하고, 두번째 실험환경은 더 활발한 것을 의미하나요??? 만일 설명을 쓰려고 한다면, specifically 어떤 환경을 말하는지를 써야 하고, 이 둘이 어떤 의미를 가지는 지를 쓰면 좋겠어요.)

We generated two sets of e-marketplace data with different parameter settings for the numbers of items, sellers and buyers. The first set consists of 1,000 items, 500 sellers, and 5,000 buyers; the second set consists of 2,000 items, 1,000 sellers, and 10,000 buyers. All simulations are based on 300 days. Each buyer is restricted to trade at most once each day.

4.2 Classification of unfair rating attacks

구매자들의 unfair한 rating은 예비소비자들이 거래 상대 (판매자)를 선택하는데 악영향을 끼칠 수 있다. 평판시스템은 구매자들의 unfair rating을 처리하여 신뢰할 수 있는 판매자 평판을 제공해야 한다. (만일 앞 문장을 쓰고 싶다면, 실험을 통해 평판시스템을 평가하기 위해서 unfair ratings를 모델한다는 것이 쓰여져야 합니다.) 본 논문에서는 구매자들의 cheating behavior에 초점을 맞춰 기존의 소개된 UR-attacker들을 다음과 같이 분류한다.

Unfair ratings of UR-attackers have undesirable and negative influence on the seller selection process of potential buyers. In our experiments, we try to evaluate whether the reputation system show a robust performance even with the existence of unfair ratings. For generating comprehensive types of UR-attacker, we classify UR-attackers based on their cheating behaviors of buyers as follows.

기존 연구들에서 다른 UR-attacker들은 총 3가지 평가 전략과 6개의 평점 패턴으로 분류할 수 있다. 평가 전략은 평가기간 내내 어떤 방식으로 평점을 매기느냐에 따라서 basic, camouflage, white washing으로 구분한다. 평점 패턴은 평점을 매길 대상 (공모 판매자 또는 경쟁 판매자)을 선정하는 방법과 구체적인 평점 생성 방법에 따라 ballot-stuffing, bad-mouthing, ballot-stuffing & bad-mouthing, r-high shifting, r-low shifting, r-high shifting & r-low shifting으

로 구분한다. 3가지 평가 전략과 6개의 평점 패턴의 조합으로 나타날 수 있는 UR-attacker의 종류는 총 18가지이다. 이에 대한 자세한 설명은 다음과 같다.

Various types of UR-attackers examined in previous research [7][10][11][12][13][15] can be classified into three evaluation schemes and six rating patterns. The combination of three evaluation schemes and six rating patterns results in 18 different types of UR-attackers. Three evaluation schemes and six rating patterns are summarized in the following.

UR-attacker들의 평가 전략은 다음의 3가지로 분류된다.

According to the evaluation strategy employed by the UR-attackers, evaluation schemes can be classified into three: *basic*, *camouflage*, *whitewashing*.

(실험을 할 때 모든 attacker들이 같은 scheme을 쓰는 거죠???)

- **Basic**: 평가기간 내내 불공평한 평점만을 매기는 단순한 (simple) 평가 전략이다 [7-15].
- **Basic**: This is the most common and simplest scheme. UR-attackers constantly grant unfair ratings—unfairly high ratings to the conspiring seller(s) and/or unfairly low ratings to rival seller(s) [7-15].
- **Camouflage**: 평판 시스템의 감시를 피하기 위해 일정기간 동안은 대다수의 일반 소비자들과 유사한 평점을 매기고, 이후에는 basic과 동일하게 불공정한 평점을 매기는 전략이다 [13].
- **Camouflage**: UR-attackers may try to camouflage themselves as normal buyers by providing some fair ratings strategically. In the camouflage scheme, UR-attackers provide fair ratings at the early stage to develop their credibility, and provide unfair ratings similar to the basic scheme at the later stage [13].
- **Whitewashing**: 한 UR-attacker가 여러 개의 사용자 계정을 두고 공모하는 전략이다. Basic과 동일하게 불공정한 평점을 매기지만 일정 기한을 두고 새로운 사용자 계정을 만들어 새 (new) 계정으로 다시 평가를 수행한다. White washing은 마치 여러 명의 일반 소비자가 평점을 매긴 것처럼 보이게 하는 전략이다 [13].
- **Whitewashing**: Most e-marketplace allows the buyer to create multiple accounts, which provides an opportunity for an UR-attacker to whitewash his low credibility by starting a new account. In the whitewashing scheme, UR-attackers unfair ratings similar to the basic scheme at the first stage and create a set of new accounts and provide unfair ratings as if many buyers would take part in ratings at the later stage [13].

UR-attacker들의 평점 패턴은 평점을 매길 대상 (공모 판매자 또는 경쟁 판매자)을 선정하

는 방법과 구체적인 평점 생성 방법에 따라 다음의 6가지로 분류된다. 6가지의 평점 패턴은 실험 결과에서 다음 순서에 따라 ①~⑥까지의 번호를 붙여 사용한다.

According to how to generate ratings and to whom the UR-attacker gives unfair ratings, rating patterns are categorized into six: *ballot-stuffing*, *bad-mouthing*, *ballot-stuffing & bad-mouthing*, *r-high shifting*, *r-low shifting*, *r-high shifting & r-low shifting*. They are given numbers from ① to ⑥ as follows.

- ① ***Ballot – stuffing***: 오직 공모 판매자들에게만 최대 (maximum) 평점을 주는 패턴이다 [7, 10, 11].

Ballot – stuffing: UR-attackers boost the conspiring sellers' reputations by providing unfairly maximum ratings (in our experiments, five) (여기서 맥시멈은 뭔가요? 원래는 high rating인데 모델의 편의를 위해 맥시멈으로 정한 것 아닌가요?) [7, 10, 11].

- ② ***Bad – mouthing***: 오직 경쟁 판매자들에게만 최소 (minimum) 평점을 주는 패턴이다 [7, 10, 11].

Bad – mouthing: UR-attackers bad-mouth the reputations of rival sellers with unfairly minimum ratings ratings (in our experiments, one) [7, 10, 11].

- ③ ***Ballot – stuffing & Bad – mouthing***: Ballot-stuffing과 Bad-mouthing을 모두 이용하는 평점 패턴이다. 즉, 공모 판매자에게는 최대 평점을 매기는 동시에 경쟁 판매자들에게는 최소 평점을 매기는 평점 패턴이다 [7, 10, 11].

Ballot – stuffing & Bad – mouthing: UR-attackers rate their conspiring sellers with maximum ratings, while their rival sellers with minimum ratings [7, 10, 11].

- ④ ***r – high shifting***: 오직 공모 판매자들에게만 평점을 매기되 일반 소비자들이 공모 판매자에게 매긴 평점의 평균 보다 r 만큼 높여서 평점을 매기는 패턴이다 [10, 12, 15].

r – high shifting: UR-attackers rate their conspiring sellers r higher than the average of ratings given by other buyers [10, 12, 15].

- ⑤ ***r – low shifting***: 오직 경쟁 판매자들에게만 평점을 매기되 일반 소비자들이 경쟁 판매자들에게 매긴 평점의 평균 보다 r 만큼 낮춰서 평점을 매기는 패턴이다 [10, 12, 15].

r – low shifting: UR-attackers rate the rival sellers r lower than the average of ratings given by other buyers [10, 12, 15].

- ⑥ ***r – high shifting & r – low shifting***: *r-low shifting*과 *r-high shifting*을 모두 이용하는 평점 패턴이다. 즉, 공모 판매자에게는 일반 소비자들의 평균 평점보다 r 만큼 높여서 평점을 매기는 동시에 경쟁 판매자에게는 일반 소비자들의 평균 평점 보다 r 만

큼 낮은 평점을 매기는 패턴이다 [10, 12, 15]. r -high shifting, r -low shifting와 r -high shifting & r -low shifting는 특정 판매자에 대한 일반 소비자들의 평점 분포를 알고 있다는 전제하에 생성 가능한 평점 패턴이다.

r – high shifting & r – low shifting: UR-attackers rate their conspiring sellers r higher than the average of ratings given by other buyers while rating the rival sellers r lower than the average of ratings given by other buyers [10, 12, 15].

Note that the three rating patterns of r -high shifting, r -low shifting, and r -high shifting & r -low shifting can be generated only when the buyers' rating distribution on a seller is known.

실험에서는 UR-attackers의 규모 (개수)에 따른 제안하는 방법의 성능을 측정한다. BSD와 RS&BSD가¹¹ UR-attack에 견고한 성능을 보이는 것을 입증하기 위해 기존의 소개된 18가지 타입의 UR-attacker들을 생성하고, 생성된 UR-attacker들을 4.1에서 생성한 e-마켓플레이스 데이터에 삽입한다.

(다른 연구에서는 18개 중 일부분만 모델하였고 우리는 이들을 체계적으로 모델하여 실험한다는 말을 explicitly 써야할 것 같습니다.) ~~In our experiments, we measure the performance of our reputation system by changing of the number of UR-attackers in e-marketplace. (이 문장이 여기 나오는 것이 어색합니다. 나중에 다시 고쳐야 할 듯.)~~

Whereas previous researches have modeled and used a small part of the eighteen types of UR-attackers, we have systematically modeled the UR-attackers by considering and analyzing their schemes and rating patterns. In order to verify that BSD and RS&BSD¹² are robust against UR-attacks, we generate the eighteen types of UR-attackers and insert them into the e-marketplace data explained in chapter 4.1.

UR-attacker들은 특정 상품 군 g 을 공격하며, 이때 UR-attacker들의 삽입 비율은 상품 군 g 에 대해 모든 일반 소비자들이 생성한 모든 평점 개수에 대한 g 에 대해 삽입될¹³ UR-attacker들이 생성할 불공평한 평점 개수로 정한다. 예를 들어, 상품 군 g 에 총 1,000개의 평점이 존재

¹¹ RS는 단독으로 사용할 경우 UR-attacker의 영향력을 줄이지 못한다. 따라서 UR-attacker가 존재하는 환경에서는 RS 단독으로 사용하지 않는다.

¹² When using RS separately from BSD, it cannot reduce the influence of UR-attackers. Therefore, we do not measure the performance of RS in an e-marketplace where UR-attackers exist.

¹³ UR-attacker도 시뮬레이션의 일반소비자와 동일하게 1일 최대 1번 거래 가능하며, 상품 구매 주기를 갖는다.

할 경우 UR-attacker의 삽입 비율을 90%로 한다는 것은 전체 거래 일수 동안 총 900개의 unfair rating를 생성할 규모의 UR-attacker를 삽입한다. 실험에서는 그 비율을 10% 에서 90% 범위에서 10% 상승시킨다. (간단하게 UR-attack의 비율이 전체 rating 중에 10~90%가 된다고 쓰지 않는 이유는 뭔가요? 이런 난해한 설명이 독자들에게 필요한가요?) 해당 상품 군 g 에서 판매자의 능력이 0.25 이하인 판매자를 공모 판매자로 간주한다¹⁴. ~~UR-attacker들은 임의의 공모 판매자와 거래를 수행한다.~~ 총 18개의 UR-attacker 각각을 따로 삽입하여 e-마켓플레이스 데이터들의 집합을 생성한다.

In experiments, we vary the ratio of UR-attacks to total ratings in the targeted item group from 10% to 90% in increments of 10%. All item groups can be targeted by a type of UR-attackers. ~~The sellers whose capability score is less than 0.25 as conspiring sellers.~~ (이것 상당히 controversial한 assumption입니다. Justify가 필요하지 않을까요?) We assume that sellers whose capability score is quite less than that of other sellers can be tempted to pump their reputations by bribing buyers. We regard the sellers whose capability score is less than 0.25 as conspiring sellers. If there are no sellers whose capability is less than 0.25 in an item group, the item group is excluded as target by UR-attackers. Like normal buyers, UR-attackers are restricted to trade at most once during a day and have a buyer's purchase cycle. Given eighteen different types of UR-attackers, we test total of thirty-six data sets.

4.3 실험 결과 Experimental results

성능 평가 metric으로 Spearman's correlation coefficient를 사용한다. 상관분석(correlation analysis)은 두 변수간에 어떤 선형적 관계를 갖고 있는 지를 분석하는 방법이다. 상관관계의 정도를 파악하는 상관계수 (correlation coefficient)는 두 변수간의 연관된 정도를 나타내는 계수로, 특히 Spearman's correlation coefficient는 데이터가 서열 척도로 구성된 경우 자료의 값 대신 순위를 이용하여 계산하는 상관계수다. 본 실험에서는 시뮬레이션상에서 정한 판매자의 능력에 따른 판매자의 순위를 정답으로 삼는다. 그리고 제안하는 방법을 사용하여 계산된 판매자간의 평판 순위가 정답과 얼마나 일치하는지를 측정한다. 두 서열척도가 완전히 일치하면 1, 상관관계가 전혀 없다면 0, 두 서열척도가 완전히 반대 방향을 띤다면 -1이 된다. 즉, 서열상관 계수가 1에 가까울수록 제안하는 방법의 성능이 우수하다고 보는 것이다.

As the performance metric, we use Spearman's rank correlation coefficient, which assesses how closely two sets of rankings agree with each other. In experiments, we measure the closeness

¹⁴ 상품 군에 판매자 능력이 0.25 이하인 판매자가 없을 경우 그 상품 군은 unfair attack 대상에서 제외한다.

between the original sellers' rankings determined by their capabilities and the rankings obtained by the proposed reputation system using Spearman's coefficient.

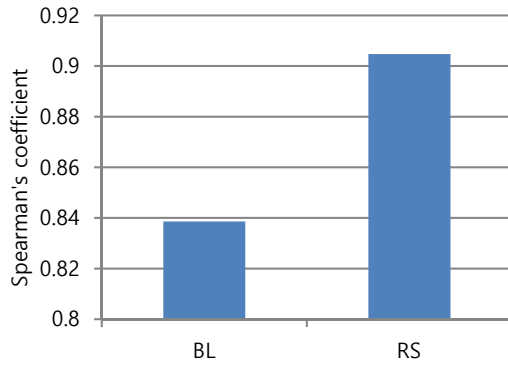
실험 결과는 크게 두 부분으로 나뉜다. 첫째, 평점 분리 기법 (RS)이 평판 계산에 유용한지를 보인다. RS의 목적은 판매자 행동에 대한 평판 계산이다. RS의 우수성을 증명하기 위해 평균을 이용한 판매자 평판 계산을 베이스라인 baseline (BL)으로 사용하여 RS과 비교 실험을 한다. 둘째, BSD와 RS&BSD가 UR-attack들에 대해 얼마나 견고한지를 기존의 방법들과의 비교를 통해 보인다. BSD의 목적은 UR-attacker들의 영향력을 줄여 평판을 계산하는 것이다. BSD의 성능 측정을 위해 총 18가지의 다양한 UR-attacker들을 생성하고, 생성된 UR-attacker들을 e-마켓 플레이스 데이터에 삽입한다. BSD의 우수성을 증명하기 위해 BSD와 기존의 UR-attacker 처리 방안인 BRS, PA, iCLUB과 성능을 비교 실험 한다. 또한 BSD와 RS을 접목시킨 평판 계산 방안인 RS&BSD의 성능도 측정하여 비교한다.

The experiments are divided into two parts. First, we investigate the usefulness of RS. In order to verify the performance of RS, we compare RS with the baseline approach, which computes the seller's reputation using simple average. Second, we investigate the robustness of BSD and RS&BSD against various UR-attacks. We compare the performances of BSD and RS&BSD with those of existing reputation systems, such as BRS, PA, and iCLUB.

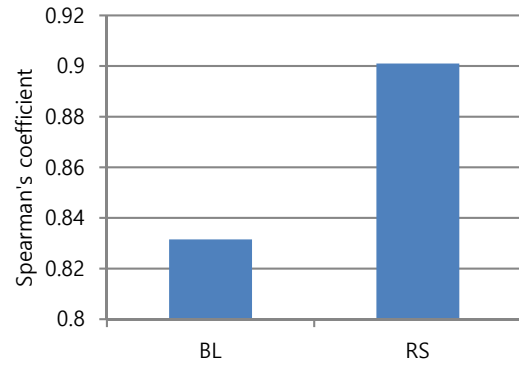
4.3.1 평점 분리 Rating Separation (RS)

RS는 단독으로 사용할 경우 UR-attacker의 영향을 줄이지 못하므로, 평점 분리 기법인 RS에 대한 평가는 UR-attacker가 존재하지 않는 상황에서 진행된다. 실험에서 사용하는 데이터는 두 개의 서로 다른 파라미터를 사용해서 생성한 시뮬레이션 데이터에서 수행된다. 본 실험에서는 평점의 평균을 이용한 평판 계산을 베이스라인 (BL)으로 사용한다. BL과 RS간의 성능 비교를 그림 4에서 확인 할 수 있다. 그림 4의 y축은 서열 상관 계수이다. 그림 4를 통해 파라미터에 관계없이 RS가 BL에 비해 좋은 성능을 보임을 알 수 있다. 이는 판매자의 평판을 계산할 때 평점을 그대로 사용하는 것보다 평점 분리를 통해 도출된 판매자의 점수만을 사용하는 것이 정확도 측면에서 보다 우수하다는 것을 의미한다.

RS, when used alone, cannot reduce the influence of UR-attackers. Therefore, we evaluate the performance of RS using two data sets where no UR-attackers exist. Figure 4 shows the comparison between the results by RS and those obtained by the baseline (BL) approach. The Y-axis in Figure 4 denotes the Spearman's correlation coefficient. RS is shown to be more accurate than BL in both data sets. The results indicate that our approach to rating separation is more effective in finding trustable seller reputations.



(a) 첫 번째 파라미터 세트
(a) Parameter set 1



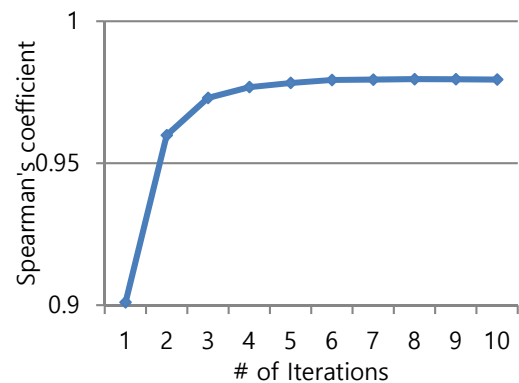
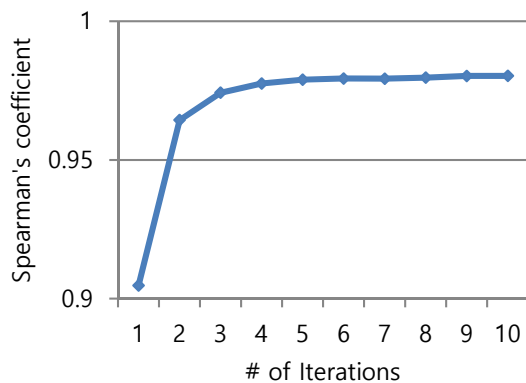
(b) 두 번째 파라미터 세트
(b) Parameter set 2

그림 4. BL과 RS의 성능 비교

Fig 4: Performance of BL and RS

그림 5에서는 RS의 반복 연산에 따른 성능의 변화를 보인다. 파라미터에 관계없이 RS은 첫 번째 반복 연산에서 정확도가 상대적으로 크게 향상되고 이후 상승폭이 서서히 줄어 수렴하는 결과를 보인다. RS을 반복해서 적용할 경우 판매자 평점 분리 과정 (SRS)에서 클러스터의 크기가 이전 평점 분리 과정보다 커질 수 있다. 그 결과 더 많은 판매자간의 상대적인 점수를 구할 수 있게 되어 보다 정확한 평판 계산이 가능해진다.

Figure 5 shows the change in performance with respect to the number of iterations in RS. The performance of the RS is improved with more iteration regardless of parameter sets. In particular, the performance is significantly improved at the second iteration, while the degree of improvement slows down with further iterations. After the fifth iteration, the performance converges to 98 percent and remains almost stable. ~~The size of clusters (which one? Both ??? Needs to be more specific) tends to increase with more number of iterations,~~ 더 많은 판매자간의 상대적인 점수를 구할 수 있게 되어 (이 말이 빠졌음) If RS works iteratively, the size of both seller and item clusters tends to increase with more number of iterations. As a result, more relative scores among the sellers and/or the items can be obtained with more sellers and/or items, which lead to higher performance.



(a) 첫 번째 파라미터 세트

(b) 두 번째 파라미터 세트

(a) Parameter set 1

(b) Parameter set 2

그림 5. BL과 RS의 정확도 변화

Fig 5: Performance changes of BL and RS

4.3.2 Handling of UR-attackers

본 절에서는 평점 신뢰도 측정 기반의 평판 계산 방법인 BSD와 평점 분리 기법을 적용한 RS&BSD의 성능을 측정한다. 비교 실험을 위해 기존의 UR-attackers 처리 방안인 BRS, PA, iCLUB을 판매자 평판 계산을 위한 방안으로 변형한다. 다음은 비교 실험을 위해 사용한 판매자 평판 계산 방법들에 대한 간략한 설명이다.

We compared the performance of BSD and RS&BSD with those of three representative reputation systems by modifying them for computing seller reputations. They are:

- **BRS**: 다수결 기법을 이용한 판매자 평판 계산 방법 [7]
- **PA**: 개인화 방법을 이용한 평판 계산 방법 [9]
- **iCLUB**: 클러스터링 기반의 평판 계산 방법 [10]
- **BRS**: Majority vote based reputation system [7]
- **PA**: Personalized reputation system [9]
- **iCLUB**: Clustering based reputation system [10]

BRS (Beta Reputation System) [7]는 한 판매자에 대한 소비자들의 평점 분포 (distribution) 분석을 통해 UR-attacker들을 검출하고, 검출된 UR-attacker의 평점을 평판 계산에서 제외하는 방법이다. BRS는 대다수 (majority)의 구매자 평점과 동떨어진 위치에 있는 평점을 매긴 소비자를 UR-attacker로 간주한다. 본 실험에서의 BRS는 특정 판매자에 대해 대다수에 비해 이질적인 평점을 매긴 소비자를 판매자의 UR-attacker로 간주하고 평판 계산에서 제외한다.

BRS (Beta Reputation System) [7] analyzes the distribution of buyers' ratings to figure out UR-attackers. In BRS, the buyers whose opinions (ratings) are not consistent with the majority are regarded as UR-attackers. BRS filters out the ratings by these buyers when computing the seller reputation.

PA (a Personalized approach) [9]는 소비자의 과거 평점 이력을 바탕으로 유사 소비자 (조언자)들을 선택하고 조언자의 평점을 참고하여 해당 판매자의 평판을 보정하는 방법이다. PA는 조언자를 선택할 때 특정 판매자에 대한 조언자의 의견이 소비자 개인과 일치하는지 확인할 뿐만 아니라 해당 조언자가 다른 판매자들에게 공정한 평가를 수행했는지도 동시에 판단하여

조언자를 최종 선택한다. PA는 조언자의 평가가 해당 판매자가 받은 대다수 (majority)의 평가와 유사할 경우 조언자는 공정한 평가를 진행한 것으로 간주한다. 본 실험에서 모든 소비자들은 자신의 조언자들의 의견을 바탕으로 판매자에 대한 새로운 평점을 갖게 되고, 각 판매자의 평판은 소비자들의 새로운 평점들에 의해 보정된다.

PA (Personalized Approach) [9] uses advisors when computing seller reputations. Advisors are selected based on two aspects: local and global similarity. To evaluate the local similarity of an advisor, PA helps a buyer to compare his ratings with the advisor's ratings given to a specific seller. The global similarity of an advisor is evaluated by comparing the advisor's ratings with the majority of the other buyers' ratings towards their commonly rated sellers. When a buyer has no experience with a seller, PA helps the buyer to get the reputation of the seller by using the global similarities of his advisors.

~~iCLUB (integrated CLUstering-Based approach) [10]은 다항 평점 (multi-nominal ratings)을 사용하는 e-마켓플레이스를 위해 고안된 클러스터링 기반의 UR-attacker 처리 전략이다. iCLUB은 클러스터링 기법인 DBSCAN을 사용하여 유사 소비자 그룹들을 형성한 후, 해당 소비자가 속하는 그룹 내의 다른 소비자 (조언자)들의 평점을 참고하여 해당 판매자의 평판을 보정하는 방법이다. iCLUB은 조언자를 선택할 때 특정 판매자에 대한 소비자와 조언자들간의 평점 유사성뿐만 아니라 해당 판매자를 제외한 다른 모든 판매자들에 대한 소비자와 조언자들간의 평점 유사성 또한 고려한다. 본 실험에서 모든 소비자들은 자신의 유사 소비자 그룹을 기반으로 판매자에 대한 새로운 평점을 갖게 되고, 각 판매자의 평판은 소비자들의 새로운 평점들에 의해 보정된다.~~

iCLUB [10] is an clustering-based approach of forming similar buyer groups by using DBSCAN[ref], which is a density-based clustering method. iCLUB helps a buyer to compute seller reputations by referring to the ratings of other buyers (called advisers) in the group to which the buyer belongs. When grouping similar buyers, iCLUB considers not only the rating similarity between a buyer and his advisors for a particular seller, but also the rating similarity between the buyer and his advisors for all other sellers. In our experiments, all buyers' ratings are adjusted by their own similar group of advisors, and then sellers' reputations are also re-computed by the buyers' adjusted ratings.

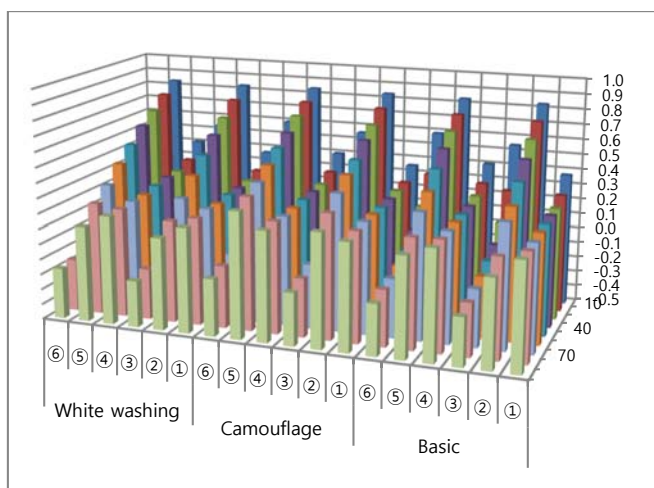


Fig. 6: Performance of BL

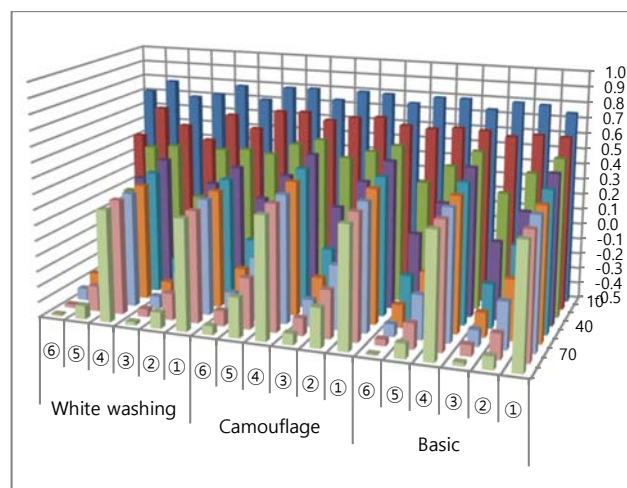


Fig. 7: Performance of BRS

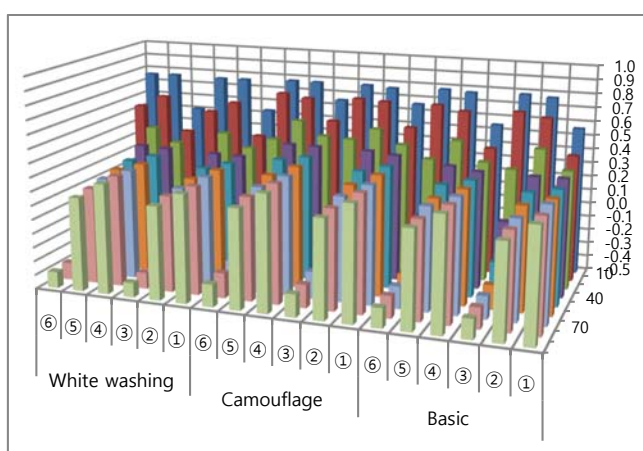


Fig. 8: Performance of PA

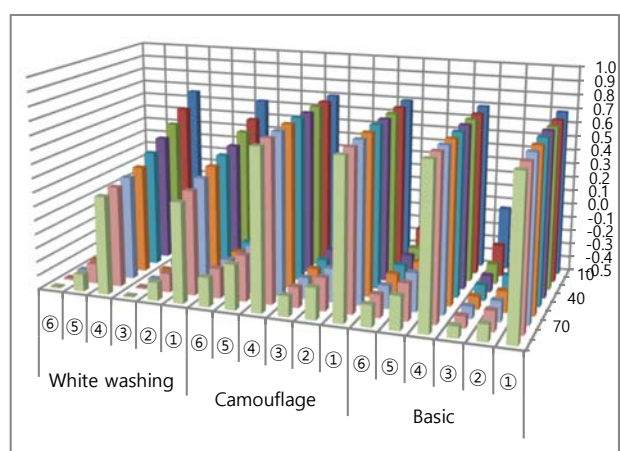


Fig. 9: Performance of iCLUB

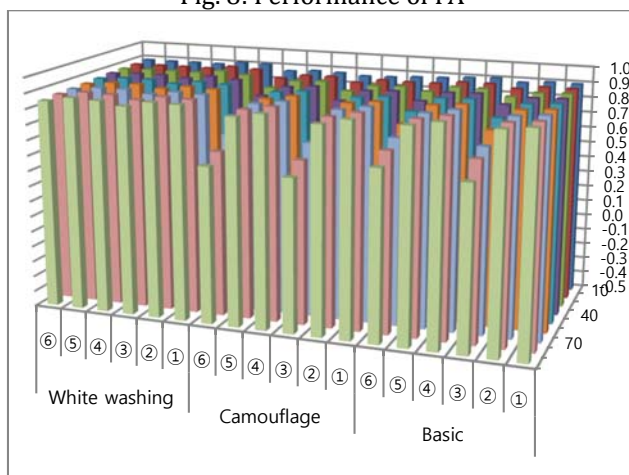


Fig. 10: Performance of BSD

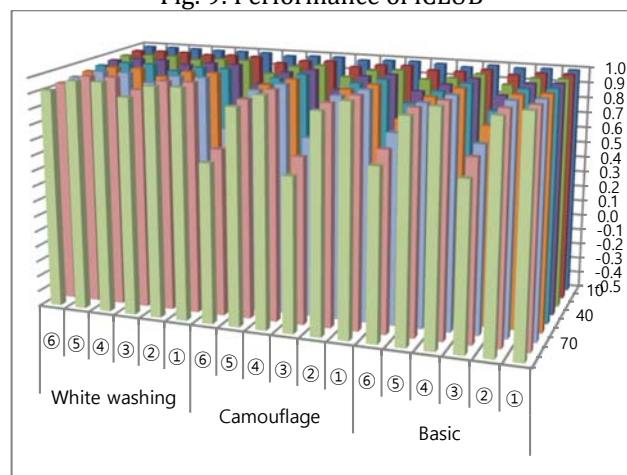


Fig. 11: Performance of RS&BSD

그림 6부터 그림 11에서 총 18개의 UR-attacker들의 공격에 대한 BL, BRA, PA, iCLUB, BSD, RS&BSD의 성능을 확인할 수 있다. 그림 6는 UR-attacks에 대한 BL의 성능이다. x축은 UR-attacker의 종류 (총 18가지), y축은 UR-attacker 삽입 비율 (10%부터 10%씩 증가하여 90%까지), z축은 서열상관계수 즉, 성능을 의미한다. x축은 오른쪽에서 왼쪽으로 순서대로 평가 전략의 경우 Basic, Camouflage, White washing 순이고 평점 패턴의 경우 Ballot-stuffing, Badmouthing Ballot-stuffing&Badmouthing, r-high shifting, r-low shifting, r-high shifting&r-low shifting 순이다. r-high shifting과 r-low shifting 평점 패턴의 경우, r 값은 0과 2사이의 값을 랜덤하게 설정하였다. BL의 경우 대부분의 UR-attack에 대해서 취약한 성능을 보임을 알 수 있다. BL은 모든 평점의 신뢰도를 동일하게 여겨 판매자의 평판을 계산하기 때문에 UR-attack에 대처하기 어렵다. 또한 UR-attacker 삽입 비율이 증가할 수록 BL의 성능이 가파르게 감소함을 알 수 있다. 그림 6의 결과를 통해 e-마켓플레이스에 UR-attacker가 존재할 때 평균을 사용하여 판매자의 평판을 계산하는 BL을 이용하면 UR-attacks의 영향력을 거의 줄이지 못하는 것을 알 수 있다.

Figures 6 to 11 show the performance of BL, BRA, PA, iCLUB, BSD, and RS&BSD, respectively, when attacked by eighteen different types of UR-attackers. In all figures, x-axis represents eighteens types UR-attackers, y-axis represents the ratio of unfair ratings generated by UR-attackers (from 10% to 90%, in the increment of 10%), and z-axis represents the Spearman's correlation coefficient. From right to left on x-axis, the three evaluation schemes are Basic, Camouflage, and White washing, and in each evaluation scheme, the six rating patterns are ① Ballot-stuffing, ② Badmouthing, ③ Ballot-stuffing& Badmouthing, ④ r-high shifting, ⑤ r-low shifting, and ⑥ r-high shifting& r-low shifting. In the case of r-high shifting and r-low shifting, the parameter r is randomly set between 0 and 2.

Figure 6 shows the performance of BL. BL performs rather poorly in most UR-attacks. Because BL is in essence the same as assigning equal trustworthiness to all ratings when computing seller reputations, it cannot handle UR-attacks. Also observed is the sharp decrease in the performance of BL with the increase of UR-attackers in the crowd. The results indicate BL which uses the average of ratings when computing the seller reputation cannot reduce the influence of UR-attackers.

그림 7는 UR-attacks에 대한 BRS의 성능이다. BRS의 경우 대다수의 구매자 평점과 동떨어진 평점을 매긴 소비자를 UR-attacker로 간주한다. 따라서 UR-attacker 삽입 비율이 증가할수록 UR-attacker들의 의견이 대다수의 의견이 되어 오히려 일반 소비자들이 UR-attacker로 간주되는 상황이 발생할 수 있다. 그림 7을 통해 할 수 있듯이 실제로 UR-attacker 삽입 비율이 증가할수록 BRS의 성능은 가파르게 떨어지며 y축이 (UR-attacker 삽입 비율) 50% 이상일 경우에는

오히려 베이스라인으로 사용하는 BL보다 열등한 성능을 보임을 알 수 있다.

Figure 7 shows the performance of BRS. BRS regards the buyers whose ratings deviates from the majority of the other buyers' ratings as UR-attackers. This strategy works when UR-attackers are a small portion of the population but falls short when the majority of buyers who rated the seller are UR-attackers. In Figure 7, the performance of BRS decreases sharply when the number of unfair ratings increases. In particular, when the ratio of unfair ratings is higher than 50%, BRS shows worse performance than BL.

그림 8은 UR-attacks에 대한 PA의 성능이다. PA 경우 역시 대부분의 UR-attacker들에 대해 취약한 성능을 보인다. PA의 경우 다른 소비자 (조언자)의 의견을 참고하여 평판을 보정한다. PA는 소비자와 유사한 평점을 매기는 소비자 또는 공정한 평가를 하는 소비자를 조언자로 선택한다. 이때 UR-attacker가 Ballot-stuffing&Badmouthing이나 r-high shifting&r-low shifting같이 복잡한 평점 패턴을 이용하였을 경우 일반 소비자와의 유사성이 높아질 수 있고, 동시에 UR-attacker 삽입 비율이 증가할수록 공정 평가의 기준이 달라지기 때문에 (PA는 대다수의 평가와 유사할 경우 공정한 평가를 진행한 것으로 간주) UR-attacker가 소비자의 조언자로 선택될 가능성이 높아진다. 그림 8의 실험 결과를 보면 실제로 PA는 Ballot-stuffing&Badmouthing나 r-high shifting&r-low shifting를 사용하는 UR-attacker에게 상대적으로 취약성을 보이고 또한 y축이 증가할 수록 그 성능이 가파르게 감소하는 것을 알 수 있다.

Figure 8 shows the performance of PA. PA is vulnerable against most UR-attacks. Especially, the performance of PA is rather poor in the following two situations. First, in the case of ③ Ballot-stuffing& Badmouthing or ⑥ r-high shifting& r-low shifting where UR-attackers use the complicated rating patterns, the ratings by UR-attackers may look similar to those of normal users. Second, when the ratio of unfair ratings of UR-attackers whose rating patterns are ③ Ballot-stuffing& Badmouthing or ⑥ r-high shifting& r-low shifting increases, the UR-attackers have more chance to be selected as advisors. Figure 8 shows that PA is vulnerable to ③ Ballot-stuffing& Badmouthing and ⑥ r-high shifting& r-low shifting and its performance sharply decrease as the ratio of UR-attackers increases.

그림 9는 UR-attacks에 대한 iCLUB의 성능이다. iCLUB은 대부분의 UR-attack에 대해 모든 UR-attackers 처리 방법 중 가장 취약한 성능을 보인다. iCLUB 또한 다른 소비자 (조언자)의 의견을 참고하여 평판을 보정한다. iCLUB은 판매자 평판 계산을 위해 DBSCAN을 이용하여 유사 소비자 그룹을 형성한다.

유사 소비자 그룹을 형성할 때 white washing 이나 camouflage 전략을 이용하는 UR-attacker가 Ballot-stuffing&Badmouthing 이나 r-high shifting&r-low shifting 복잡한 평점 패턴

을 이용하였을 경우 일반 소비자로 간주되어 일반 소비자 그룹에 UR-attacker들이 꽤 삽입되었을 거라 예상된다. 그림 9의 실험 결과를 보면 iCLUB은 white washing 전략을 사용한 UR-attacker들 특히 그 중에서도 Ballot-stuffing&Badmouthing 이나 r-high shifting&r-low shifting 평점 패턴을 사용한 UR-attacker들의 공격에 매우 취약한 성능을 보이는 것을 알 수 있다.

Figure 9 shows the performance of iCLUB, which is the worst among those compared. Note that the DBSACN in iCLUB uses a cosine similarity measure, and the parameters are set as follows: radius = 10 and MinPts = 18. In particular, iCLUB is vulnerable when UR-attackers adopt evaluation schemes of white washing or camouflage and also use the complicated rating patterns of ③ Ballot-stuffing& Badmouthing or ⑥ r-high shifting& r-low shifting. It is because the ratings of UR-attackers are more likely to be regarded as those of normal users in these cases that a considerable number of them are selected as advisors.

그림 10는 UR-attacks에 대한 BSD의 성능이다. 제안하는 BSD는 대부분의 UR-attacker에 대해 매우 견고한 성능을 보이며 특히 기존 방법인 BRS, PA, iCLUB들이 매우 취약한 성능을 보이는 white washing 평가 전략을 사용하며, Ballot-stuffing&Badmouthing 이나 r-high shifting&r-low shifting 평점 패턴을 사용하는 UR-attacker 영향을 거의 받지 않는 것을 알 수 있다. 상대적으로 camouflage 방식의 UR-attacker들에게 취약한 성능을 보이지만 BL을 비롯한 기존 방법들과 비교해 보았을 때 y축 전 구간에서 월등하게 좋은 성능을 보인다. 이를 통해 소비자의 활동성, 다양성, 대중성이 판매자 평판 보정에 유용함을 알 수 있다.

Figure 10 shows the performance of BSD. BSD performs better than the other reputation systems compared regardless of the ratio of UR-attackers in the system. In particular, compared to BRS, PA, and iCLUB, BSD is very robust against the UR-attackers adopting white washing and using the rating patterns of ③ Ballot-stuffing& Badmouthing or ⑥ r-high shifting& r-low shifting. BSD is relatively vulnerable to the case with camouflaged UR-attackers because they provide fair ratings at the early stage which can be assigned higher weights when computing seller reputations. When compared to other reputation systems, however, BSD shows the good performance under all types of UR-attacks. The good performance of BSD confirms that computing the seller reputations by identifying and exploiting the three key factors, activity, diversity, and universality brings out a reliable reputation system which is robust against any types of UR-attacks.

그림 11는 UR-attacks에 대한 RS&BSD의 성능이다. RS&BSD는 BSD에 비해 모든 UR-attacks에 대해 전반적으로 조금 더 좋은 성능을 보인다. 이를 통해 평점 분리 기법인 RS을 BSD에 접목하면 UR-attacks의 영향력을 줄이는데 도움이 되는 것을 알 수 있다.

Figure 11 shows the performance of RS&BSD, which is even better than BSD. The result confirms that the integration of RS into BSD can attain more trustable seller reputations by using the seller scores only and by reducing the influence of UR-attackers even further.

표 2. UR-attack에 대한 평판 계산 방법들의 성능 (평가 전략 별)

Table 2. Performance of reputation systems (with respect to evaluation schemes)

Evaluation schemes	BL	BRS	PA	iClub	BSD	RS&BSD
Basic	0.262	0.182	0.252	0.040	0.848	0.922
Camouflage	0.264	0.182	0.262	0.040	0.845	0.921
White washing	0.276	0.198	0.269	0.051	0.840	0.916
Average	0.267	0.187	0.261	0.044	0.844	0.920

표 3. UR-attack에 대한 평판 계산 방법들의 성능 (평점 패턴 별)

Table 3. Performance of reputation systems (with respect to rating patterns)

Rating patterns	BL	BRS	PA	iClub	BSD	RS&BSD
① Ballot-stuffing	0.242	0.422	0.366	0.566	0.886	0.988
② Badmouthing	0.492	0.145	0.375	-0.271	0.879	0.972
③ Ballot-stuffing& Badmouthing	0.048	-0.021	0.005	-0.354	0.782	0.847
④ r-high shifting	0.260	0.422	0.371	0.570	0.884	0.988
⑤ r-low shifting	0.523	0.147	0.385	-0.205	0.874	0.959
⑥ r-high shifting& r-low shifting	0.083	-0.026	0.012	-0.303	0.786	0.856
Average	0.275	0.181	0.252	0.001	0.849	0.935

표 2와 표 3에서는 본 논문에서 사용된 모든 UR-attack에 대한 모든 평판 계산 방법들의 성능을 종합적으로 파악 할 수 있다. 표 2는 UR-attacker의 각 평가 전략에 대한 모든 평판 계산 방법들의 성능이다. 표 2에서 평판 계산 방법의 성능 (표의 각 셀이 나타내는 값)은 각 평가 전략 (Basic, Camouflage, 또는 White washing)을 사용하는 모든 UR-attacker (6개의 평점 패턴 모두)를 대상으로 총 9번의 UR-attacker 삽입 (10%부터 10% 증가해서 90%까지) 상황에서 각 평판 계산 방법 성능들의 평균 (총 54개 실험 결과의 평균)으로 나타낸다. 표 3은 UR-attacker의 각 평점 패턴에 대한 모든 평판 계산 방법들의 성능이다. 표 3에서 평판 계산 방법의 성능 (표의 각 셀이 나타내는 값)은 각 평점 패턴을 사용하는 모든 UR-attacker (3개의 평가

전략 모두)를 대상으로 총 9번의 UR-attacker 삽입 (10%부터 10% 증가해서 90%까지) 상황에서
서의 각 평판 계산 방법 성능들의 평균 (총 27개 실험 결과의 평균)으로 나타낸다.

표 2와 3을 통해 제안하는 방법들은 (BSD와 RS&BSD) 어떠한 평가 전략과 평점 패턴에도
상대적으로 매우 견고한 성능을 보이는 것을 알 수 있다. 특히 RS기법을 접목하는 RS&BSD의
경우 r-high shifting&r-low shifting 평점 패턴을 사용하는 UR-attack에 대해서는 평균적으로
0.85 이상의 성능을 그 외의 모든 UR-attack에 대해 평균적으로 0.9 이상의 견고한 성능을 보
임을 알 수 있다.

Tables 2 and 3 summarize the performance of reputation systems with respect to evaluation
schemes and rating patterns, respectively. Each cell in Table 2 captures the average of
performance across eighteen rating patterns and nine ratios of unfair ratings (from 10% to 90%,
with the increment of 10%). Each cell in Table 3 depicts the average of performance across three
evaluation schemes and nine ratios of unfair ratings. As shown in Tables 2 and 3, BSD and
RS&BSD are very robust against any UR-attacks regardless of evaluation schemes and rating
patterns. In particular, while the performance of BL, BRA, PA, and iCLUB is less than 0.3 on average,
the performance of BRS is greater than 0.84 on average against all types of UR-attacks.
Furthermore, RS&BSD achieves about 10% better performance than BSD showing that its
performance is greater than 0.92 on average.

5. 결론

E-marketplace내 구매자 평점에는 판매자에 대한 평가와 상품에 대한 평가가 혼재되어 있
으며, 구매자 평점을 단순히 집계하여 평판을 계산하면 문제가 되는 상황들이 발생한다. 본 논
문에서는 신뢰할 수 있는 판매자 평판 계산을 위해, 평점으로부터 판매자에 대한 평가만을 추
출하는 평점 분리 기법 RS와 소비자의 활동성, 다양성, 대중성 측정을 통해 평점의 신뢰도를
계산하는 방안 BSD을 고안했다. 또한 RS와 BSD를 접목하여 이용하는 평판 계산 방법인
RS&BSD를 제안했다. 제안하는 방법들의 성능 평가를 위해 필요한 데이터 생성을 위해 기존의
단순한 시뮬레이션에서 벗어난 e-마켓 플레이스의 현실성을 반영한 고차원 시뮬레이션을 구축
하였다. 이후 생성된 데이터에 기존 문헌에서 소개한 총 18개의 UR-attacker들을 삽입하여 실험
데이터를 생성하였다. RS&BSD는 다양한 UR-attack에 대해서 기존의 평판 계산 방법인 BRS,
PA, iCLUB에 비해 월등하게 견고한 성능을 보임을 실험을 통해 증명했다. BRS, PA, iCLUB는 상
당수의 UR-attacker에게 취약한 성능을 보인 반면에 RS&BSD는 대부분의 UR-attack에 견고한
성능을 보였다.

Two observations regarding the seller reputation have led us to the development of RS&BSD.
First, the buyer's rating in e-marketplace is a compound evaluation of the capability of a seller

and the quality of an item. Rating Separation (RS), proposed in this paper, separates the rating into 'seller score' and 'item score' and computes the seller reputation using the seller scores only. Second, there exist unfair ratings coming from various types of cheating behaviors between buyers and sellers in e-marketplace. Better Seller Discovery (BSD), also proposed in this paper, evaluates the level of trustworthiness of ratings based on activity, diversity, and universality and computes the seller reputation by adjusting the original reputation based on the level of trustworthiness of ratings. In experiments, the proposed reputation system which is RS&BSD integrating both RS and BSD shows the best performance compared with existing reputation systems, BRS, PA and iCLUB. The existing reputation systems are vulnerable to most UR-attackers. On the other hands, RS&BSD is very robust against all kinds of UR-attackers.

6. 참고문헌

- [1] <http://en.wikipedia.org/wiki/EMarketPlace>
- [2] A. Jøsang, R. Ismail, and C. Boyd. "A survey of trust and reputation systems for online service provision," *Decision Support Systems*, Vol. 43, Issue 2, pp. 618-644, 2007.
- [3] A. M. Kramer, "Self-selection Bias in Reputation Systems," *Trust Management*, Vol. 238, pp. 255-268, 2007.
- [4] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood, "The Value of Reputation on eBay: A Controlled Experiment," *Experimental Economics*, Vol. 9, Issue 2, pp. 79-101, 2006.
- [5] L. Cabral and A. Hortacsu, "The Dynamics of Seller Reputation: Theory and Evidence from eBay," *The Journal of Industrial Economics*, Vol. 58, Issue 1, pp. 54-78, 2010.
- [6] D. Houser and J. Wooders, "Reputation in Auctions: Theory, and Evidence from eBay," *Journal of Economics & Management Strategy*, Vol. 15, Issue 2, pp. 353-369, 2006.
- [7] A. Whitby, A. Jøsang and J. Indulska, "Filtering Out Unfair Ratings in Bayesian Reputation Systems," *The Icfa Journal of Management Research*, Vol. 4, No. 2, pp. 48-64, 2005.
- [8] W. T. L. Teacy, J. Patel, N. R. Jennings, and M. Luck, "Travos: Trust and Reputation in the Context of Inaccurate Information Sources," *Autonomous Agents and Multi-Agent Systems*, Vol. 12, No. 2, pp. 183-198, 2006.
- [9] J. Zhang and R. Cohen, "Evaluating the Trustworthiness of Advice about Seller Agents in E-marketplaces: A Personalized Approach," *Electronic Commerce Research and Applications*, Vol. 7, No. 3, pp. 330-340, 2008.
- [10] S. Liu et al., "iCLUB: an Integrated Clustering-based Approach to Improve the Robustness of Reputation Systems," In *Proc. of the 10th Int'l Joint Conf. on Autonomous Agents and Multiagent Systems*, pp. 1151-1152, 2011.
- [11] C. Dellarocas, "Immunizing Online Reputation Reporting Systems against Unfair Ratings and Discriminatory Behavior," In *Proc. of the 2nd ACM conf. on Electronic commerce*, pp. 150-157, 2000.
- [12] A. Jsang and R. Ismail, "The Beta Reputation System," In *Proc. of the 15th bled electronic commerce conf.*, pp. 41-55, 2002.
- [13] L. Zhang, s. Jiang, J. Zhang, and W. K. Ng, "Robustness of Trust Models and Combinations for Handling Unfair Ratings," *Trust Management VI*, Vol. 374, pp. 36-51, 2012.

- [14] H. Fang, J. Zhang, M. Şensoy, N. M. Thalmann, "SARC: Subjectivity Alignment for Reputation Computation," In *Proc. of the 11th Int'l Conf. on Autonomous Agents and Multiagent Systems*, pp. 1365-1366, 2012.
- [15] K. Regan, P. Poupart, and R. Cohen, "Bayesian Reputation Modeling in E-marketplaces Sensitive to Subjectivity, Deception and Change," In *Proc. of the National Conf. on Artificial Intelligence*, 2006.
- [16] E. Lim, V. Nguyen, N. Jindal, B. Liu, and H. Lauw, "Detecting Product Review Spammers Using Rating Behaviors," In *Proc. of ACM International Conference on Information and Knowledge Management*, pp. 939-948, 2010.
- [17] H.-K. Oh, S.-W. Kim, and S. Park, "Trustable Aggregation of Online Ratings," In *Proc. of ACM Int'l Conf. on Information and Knowledge Management*, pp. 1233-1236, 2013.
- [18] M. Chen and J. P. Singh, "Computing and Using Reputations for Internet Ratings," In *Proc. of the 3rd ACM conf. on Electronic Commerce*, pp. 154-162, 2001.
- [19] A. Irissappane, S. Jiang, and J. Zhang, "Towards a Comprehensive Testbed to Evaluate the Robustness of Reputation Systems Against Unfair Rating Attack," *User Modeling Adaptation and Personalization Workshops*, 2012.
- [20] P. Chandrasekaran and B. Esfandiari, "A Model for a Testbed for evaluation reputation system," In *Proc. of Int'l Conf. on Trust, Security and Privacy in Computing and Communications*, pp. 296-303, 2011.
- [21] R. Kerr and R. Cohen, "TREET: The Trust and Reputation Experimentation and Evaluation Testbed," *Electronic Commerce Research*, Vol. 10, Issue 3, pp. 271-290, 2010.
- [22] J. Howe, *The Rise of Crowdsourcing*, Wired, 2006.
- [23] C. Leadbeater, *WE-THINK: Mass Innovation, not Mass Production*, Profile Books, 2008.
- [24] P. Levy, *L'Intelligence Collective: Pour Une Anthropologie Du Cyberspace*, La Découverte, 1997.