SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

May 2019, Vol. 62 052101:1–052101:17
https://doi.org/10.1007/s11432-018-9511-1

A trust-aware random walk model for return propensity estimation and consumer anomaly scoring in online shopping

Xiaolin LI[1], Yuan ZHUANG[1], Yanjie FU[2*] & Xiangdong HE[3*]

[1]*School of Business, Nanjing University, Nanjing* 210093*, China;*
[2]*Department of Computer Science, Missouri University of Science and Technology, Rolla* MO 65401*, USA;*
[3]*Network and Information Center, Nanjing University, Nanjing* 210093*, China*

Received 23 April 2018/Revised 1 June 2018/Accepted 25 June 2018/Published online 19 March 2019

Abstract In online shopping, most of consumers will not clear their return reasons when submitting return requests (e.g., select the option "other reasons"). Prior literature mostly investigates into the return event at the transaction level, and the underlying force of returns remains untracked. To deal with this problem, we propose a machine learning algorithm named as trust-aware random walk model (TARW). In the proposed model, four patterns of consumers can be identified in terms of return forces: (i) selfish consumers, (ii) honest consumers, (iii) fraud consumers, and (iv) irrelevant consumers. To profile consumers' return patterns, we capture consumers' similarities in order preferences and return tendencies separately. Based on consumers' similarities, we obtain a return pattern trust network by introducing the trust network and collaborative filtering algorithms. Subsequently, we develop two important applications based on the trust network: (i) estimating consumers' return propensities for product types; (ii) scoring the anomaly for consumers' returns for one product. Finally, we conduct extensive experiments with the real-world data to validate the model's effectiveness in predicting and tracing consumers' returns. With the proposed model, we can help retailers improve the conversion rates of selfish consumers, retain honest consumers, and block fraud consumers.

Keywords machine learning, return abuse, random walk, collaborative filtering, return pattern

Citation Li X L, Zhuang Y, Fu Y J, et al. A trust-aware random walk model for return propensity estimation and consumer anomaly scoring in online shopping. Sci China Inf Sci, 2019, 62(5): 052101, https://doi.org/10.1007/s11432-018-9511-1

1 Introduction

Given the rise in popularity of the online shopping, liberal product return policies (i.e., no-hassle return policies) have been widely adopted by online retailers to promote sales, which also bring on serious return abuses. According to the CIConsulting[1)], in 2016, the average online return rate of clothing products is up to 30% in China. Besides, at present, existing systems can only partially prevent malicious returns through the dispute process, and the financial penalties for consumers' return abuses are relatively small. Therefore, it is needed for online retailers to explore an effective way for detecting and restricting return abuses. To this end, we need to capture the patterns of consumers' return events (e.g., malicious or non-malicious). In other words, it is crucial for retailers to understand underlying forces of the return events, and ultimately to identify and prevent the return abuse events by making malicious return consumers predictable and traceable.

* Corresponding author (email: fuyan@mst.edu, hexd@nju.edu.cn)
1) http://www.ocn.com.cn/hongguan/201611/zcsyo13120558-2.shtml.

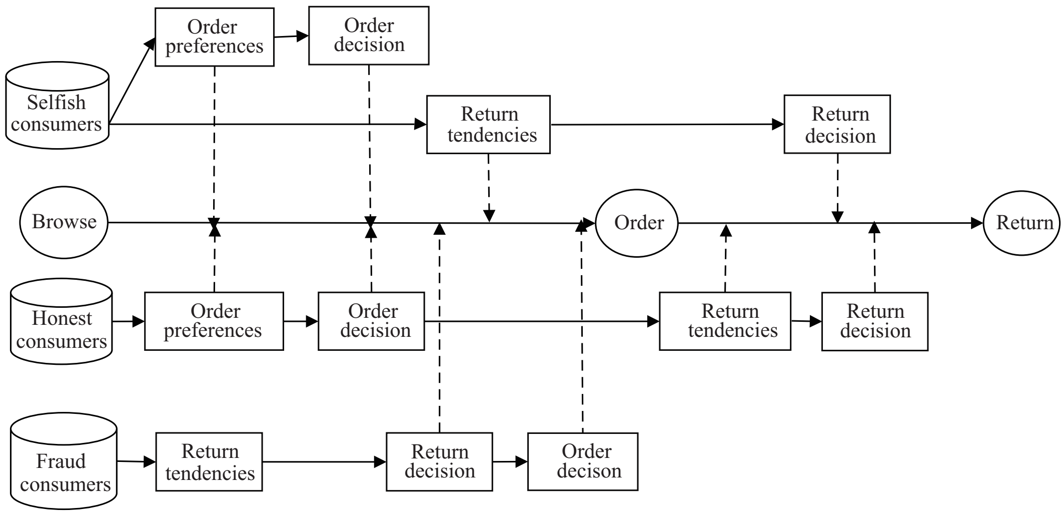

Figure 1 Purchase decision process with respect to three consumer groups.

However, it is traditionally challenging for online retailers to find out the underlying forces of consumers' returns. In practice, most of consumers will not clear their return reasons when submitting return requests (e.g., select the option "other reasons" as the return reason). In prior literatures, an individual return event is mostly associated with one consumer and one transaction [1], thus the purpose of the return event remains untracked. Therefore, in this paper, we propose to gain insights into the return abuses at the level of consumers and product types, and then capture the return patterns of consumers. To extend the analysis of the return patterns, the following three key questions need to be answered:

- Research question 1: Can consumers be segmented into several latent groups in terms of their return patterns?
- Research question 2: How does the return pattern of different consumer groups impact the return decision process of products?
- Research question 3: How can the return propensities and anomaly scores of consumers be strategically estimated?

To deal with the first research question, we propose to distinguish different return patterns in terms of the underlying return forces. Generally, there are four types of return forces in practice [2]: (i) expectation gaps (i.e., the actual product is not in line with the customer's preference or expectation) [3]; (ii) retailers' issues (e.g., product quality problem, inventory problem, etc.); (iii) return frauds; and (iv) consumers' accidental mistakes (e.g., order the wrong product, etc.). In terms of return forces, consumers can be categorized as four groups or four patterns: (i) selfish consumers; (ii) honest consumers; (iii) fraud consumers; and (iv) irrelevant consumers.

As for the second question, we decompose the consumer purchase process into two stages: (i) the order stage and (ii) the return stage, according to the work in [4]. We depict how consumer return patterns impact the return decision process in Figure 1. Specifically, honest consumers usually return a product only when the product has quality issues. Selfish consumers indeed have return tendencies for one product before they place orders. If the trial experience is lower than their expectations, the product may be returned. And fraud consumers usually do not have any preferences and need to purchase a product. Before ordering the product, fraud consumers clearly know that they will return the product, and they just request free trials, the refund or even extra compensation. In summary, selfish consumers are those who abuse liberal product return policies without malicious intentions; fraud consumers are not just liberal product return policies abusers but also malicious users. To help retailers, we propose to improve the conversion rates of selfish consumers, retain honest consumers, and block fraud consumers.

Along these lines, we can deal with the third research question by quantifying consumers' return patterns with consumer order preferences and return tendencies, as shown in Figure 2. In particular,

- Selfish consumers have certain preferences to order a product, and appear high return tendencies on

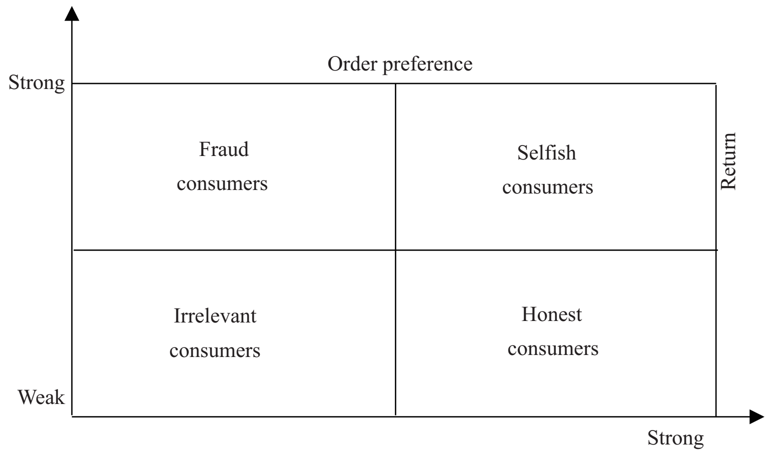

Figure 2 Four groups of consumers in online e-commerce.

the product.

• Honest consumers have clear need and preferences to order a product, and mostly will not return the product if there are no quality issues.

• Fraud consumers have no need and preferences to purchase the product, but they intentionally order it for free trials, or just to return the product and claim for compensations.

• Irrelevant consumers have no need and preferences to order a product and thereby have no return tendencies on the product.

In this paper, we only take the first three consumer groups into consideration. Besides, we study large-scale product return events in an Alibaba B2C platform, and find that consumers' individual differences are proved to be more potential for understanding and profiling the heterogeneity of return behaviors. Hence, we propose to identify the return patterns by segmenting consumers with a machine learning algorithm. The proposed model is named as trust-aware random walk model (TARW). Specifically, we first introduce the idea of the trust network in the TrustWalker algorithm, and then transform order events records as the initial version of the trust network. According to the framework of the RWR model, we further use return events records to generate two types (product return probabilities and consumer return tendencies) of consumer trust relations. After that, we adjust the initial consumer trust network with these trust relations to develop the final consumer trust network. Finally, we obtain the consumer trust network that can be used to segment the three groups of consumers in terms of return patterns. Each group of consumers for a product type is indeed a connected community in this network and can be identified by their unique network-related properties. For instance, given a product, if a consumer is not trusted by other consumers ordering the same product in this network; that is this consumer is different from other consumers with respect to both order preferences and return tendencies, then this consumer is prone to be a fraud consumer.

The consumer trust network are used to enable two important applications:

• Estimating the return propensity of a consumer for a product type;

• Scoring the anomaly for returns and identifying the return pattern for each pair of a consumer and a product type.

In this paper, we conduct extensive experiments with the real-world data. The experimental results validate the effectiveness of the proposed model in predicting return propensities, estimating anomaly scores, and ultimately categorizing consumers in terms of return patterns. With the proposed model, we can help retailers take appropriate measures to control the risk and loss of the return abuses.

2 Research background

The related work can be categorized as (i) malicious behavior detection and (ii) random walk algorithms.

2.1 Malicious behavior detection

To protect consumers, retailer credit assessment has received significant focus in prior work [5, 6], while less attention has been given to consumer malicious behavior detection. In e-commerce, retailers always detect malicious consumers by assessing consumers' credit scores. And the credit assessment of one entity is defined as the process of constructing large-scale word-of-mouth networks by collecting, disseminating and aggregating entities' public information [7]. The work in [8] took into account the return rate factor when constructing the credit scoring model. And the work in [1] introduced the ideas of the collaborative filtering and the latent factor model to predict consumers' return propensities.

In other relevant sectors, such as the banking, health care, and insurance, statistics or machine learning algorithms are widely adopted to predict the propensity of consumer malicious behaviors, such as support vector machines, decision trees, neural networks [9–11]. The work in [12] classified normal and abnormal ECG records based on the lead convolutional neural network. However, in these algorithms, relations or interactions between consumers and different products are not taken into account. Hence, it is difficult for retailers to detect consumers' specific malicious behaviors, such as the return frauds. The work in [13] considered differences in the sharing habits of users to reduce interference in multiple user sharing behaviors and malicious fraud in P2P environments.

2.2 Random walk algorithm

The core idea of the collaborative filtering is to capture relations between users and items based on the rating records, and then make according recommendations. Relevant algorithms can be divided into two groups: (i) memory-based models, such as user-based algorithms [14] and item-based algorithms [15]; (ii) model-based models, such as Bayesian networks [16, 17], latent factor models [18], and bipartite graph models [19, 20]. In [21], a trust-based algorithm is introduced to assess the consumer credit scores.

Among these models, the random walk algorithm transforms the rating records as one user-item bipartite graph [22] and performs well with the interpretability, which is an application of the Markov network. Edges connecting users and items are weighted to profile the strength of relations between users and items. The topic-sensitive random walk algorithm assigned topic weights to the edges [23] and significantly improved the performances of recommendations. The work in [20] used a Markov decision process (MDP) to model the passenger seeking process and found the best move for a vacant taxi. In the anomaly detection domain, the random walk with restart algorithm (RWR) is developed to generate a user similarity matrix. In [24], one user is considered as abnormal for an item if the user is not similar with other users who related with the same item. In collaborative filtering, relevant algorithms mostly adopted a trust network approach, such as TrustWalker [25], SimRank [26], and project tag-oriented random walk. To overcome the challenge of the data sparseness, these algorithms brought in social networks to identify trusted users. Prior work mainly focused on optimizing the trust network by developing more effective weight metrics. For instance, Zhang et al. [27] introduced the social tagging information. Alexandridis et al. [28] utilized one probability distribution to model the user-user and user-item relations. The work in [29] integrated a weighted average method (WAM) into the random walk (RW) framework to employ social ties, behavior context, and personal information.

In this paper, we introduce the idea of the trust network and develop the TARW to profile the patterns of consumers' returns, based on a RWR framework.

3 Preliminary analysis

The data set used to calibrate and validate the model is extracted from the purchase records of one online cosmetics retailer in Taobao, the largest B2C platform in China. As we can see in Table 1, this data set consists of 33558 purchase records generated by 6223 consumers and 990 products from January 12, 2013 to January 12, 2014. Table 2 shows the statistics of important attributes in this data set. In this section, we propose to identify the key factors that related to the amount difference of product orders

Table 1 Important statistics of purchase records

Data sources	Description	Statistics
Customer	Number of consumers	6223
Product	Number of products	990
Order	Number of orders	143835
	Average orders per consumer	23.11
	Average orders per product	145.29
Return and refund	Number of returns or refunds	9886
	Average returns or refunds per consumer	1.59
	Average returns or refunds per product	9.99

Table 2 Statistics of important attributes

Attribute	Mean	Max	Min	Standard deviation
Customer_credit	389.3868	25471	0	547.8237
Active_time	1412.237	3819.027	0	749.3843
Consumer_order_num	23.11345	78000	1	989.2386
Consumer_return_num	1.588623	507	0	7.700736
Item_price	81.85295	3050	0.1	113.0857
Exist_time	580.5816	722.1351	4.003935	194.6978
Has_warranty	0.1252458	1	0	0.3310024
Has_invoice	0.0021157	1	0	0.0459492
Has_showcase	0.3703141	1	0	0.482896
Sub_stock	1.053251	2	1	0.2245371
Has_discount	0.9733596	1	0	0.1610326
Discount_fee	17.38692	12000	0	93.27477
Post_fee	3.767954	5	0	2.15463
Trade_time	6.515412	190.7472	0	4.918788
Payment	94.29233	400000	0	3019.021

Table 3 Attributes description

Feature	Attribute	Description
Consumer profile	Buyer_order	Number of consumer's orders
	Buyer_credit	Consumer's credit value
	Active_time	Consumer's active time
Product profile	Exist_time	Product's list time
	Item_price	Product's price
	Sales_num	Product's sales volume
Transaction profile	Payment	Purchase's payment
	Trade_time	Purchase's time
	Discount_fee	Purchase's discount

and returns. To this end, we calculate an information gain ratio for each attribute in the extracted data set, which measures how important an attribute is to the target value.

Table 3 lists the top-9 attributes based on information gain ratio. These attributes can be divided into three groups: (i) the product profile, (ii) the transaction profile (marketing profile), and (iii) the consumer profile. We further diagram information gain ratios of the top-9 attributes in Figure 3, where the attribute "payment" and the attribute "trade_time" are removed for their direct causality with order or return decisions. As shown in Figure 3, attributes in the transaction profile and the product profile are more important to the order decision ("order_num") than the return decision ("return_num"). On the contrary, attributes in the consumer profile are more important to the return decision than the order decision. Besides, the consumer profile is more important than the transaction profile and product profile to the return decision.

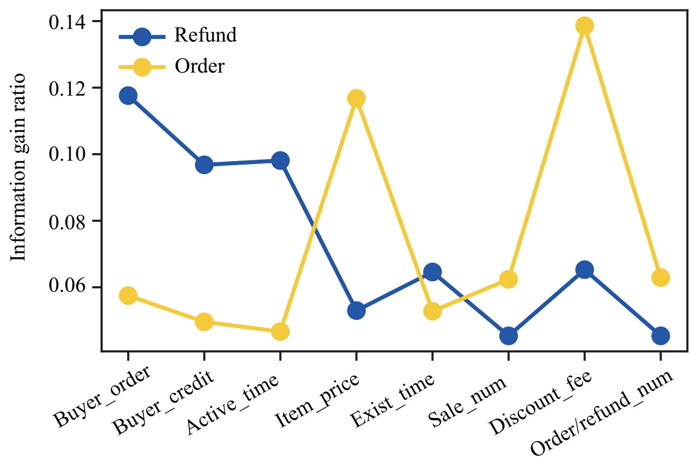

Figure 3 (Color online) Comparison of information gain ratio.

In summary, we can infer that the consumer profile is a more crucial factor related to the difference in the amount of product returns than the product profile and transaction profile, especially based on similar order preferences. And the consumer profile is considered as one's relatively stable psychological tendencies and characteristics, which can lead to one's consistent and sustained reflection to the surroundings [30]. Along these lines, it is more effective to categorize consumers to understand the heterogeneity in return events. In this paper, we propose to quantify return patterns at the level of consumers by generating a consumer trust network. Moreover, we will estimate consumers' return propensities and exploit returns' anomaly scores for each product by further extending the trust network at the level of product types.

4 TARW model

In this paper, we develop the TARW by generating trust relations and a trust network among consumers.

4.1 Problem statement

In this paper, four intuitions are proposed as follows:

Intuition 1. The trust from consumer u to v is defined as the probability that consumer u will trust and adopt consumer v's choice when making decisions (i.e., order decision, return decision) among all the consumers.

Intuition 2. There are three groups of consumers to be categorized in terms of return patterns in this paper: (i) selfish consumers, (ii) honest consumers, and (iii) fraud consumers, and these three groups can be quantified and profiled with respect to consumer order preferences and return tendencies to products.

Intuition 3. It is effective to conduct consumer categorization for understanding the heterogeneity in return events, that is, consumers in the same group always appear similar return tendencies and propensities for one product, otherwise not.

Intuition 4. For one product, consumers in the same group are trusted or distrusted by similar consumers based on return patterns, otherwise not.

Based on the four intuitions, we leverage the collaborative filtering and the trust network to develop the TARW. Main notations used in this paper are shown in Table 4. Specifically, as we can see in Figure 4, we generate a consumer trust network with two steps. First, we develop an enhanced Pearson metric to construct three similarity graphs: (i) consumer order similarity graph (SimCO), (ii) consumer return similarity graph (SimCR), and (iii) product return similarity graph (SimPR), based on order events and past corresponding return events. Second, we develop the TARW based on the three graphs. In particular, we first exploit the SimCO to develop an initial version of a consumer trust network (Intuitions 1 and 2). Then, for each iteration of the random walk, we adjust the weights of the edges in this network

Table 4 Symbol description

Notation	Description
u,v	Consumer u,v $\in\{0,\ldots,m-1\}$
i,j	Product i,j $\in\{0,\ldots,n-1\}$
SimCO	The n-by-n consumer order similarity graph
SimCR	The n-by-n consumer return similarity graph
SimPR	The m-by-m product return similarity graph
k	The depth of the walk at the moment
t_0	The initial version of consumer trust network
ρ	Consumer trust network
ω	Consumer return trust relations
α	The stopping probability
RP_u	The return products of the consumer u
$\hat{r}$	Predicted return propensity
r	Return propensity
mutual_trust	The mutual-trust between two consumers
as	The anomaly score

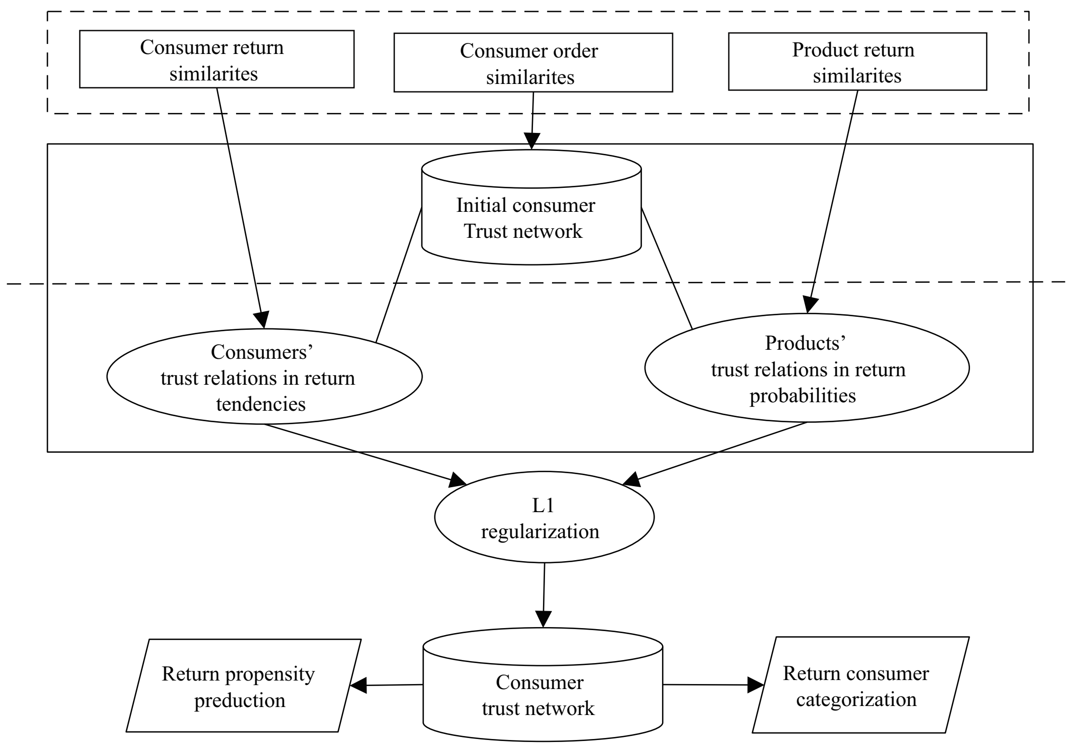

Figure 4 The general flow of proposed method.

with the trust relations extracted from SimCR and SimPR. The final trust network is used to enable two important applications presented in Figure 4: (i) estimating the return propensity of a consumer for a product type (Intuition 3); (ii) scoring the anomaly for returns and identifying the return pattern for each pair of a consumer and a product type (Intuitions 2 and 4).

The three similarity graphs are constructed based on an enhanced Pearson metric in the proposed model (TAWR). The enhanced Pearson similarity $\mathrm{sim}(X,Y)$ between two normalized vectors (normalized by $l1$ norm in (1)) X and Y is given by (3):

$$X=\frac{X}{\|X\|_1},\qquad Y=\frac{Y}{\|Y\|_1}, \tag{1}$$

$$\mathrm{corr}(X,Y)=\frac{\sum XY+\frac{\sum X\sum Y}{N}}{\sqrt{(\sum X^2-\frac{(\sum X)^2}{N})(\sum Y^2-\frac{(\sum Y)^2}{N})}},\tag{2}$$

$$\mathrm{sim}(X,Y)=\frac{\mathrm{corr}(X,Y)}{1+\mathrm{e}^{-\frac{|U_{X,Y}|}{2}}},\tag{3}$$

where $\mathrm{corr}(X,Y)$ is the standard Pearson metric, N is the length of X and Y. In this paper, we introduce $U_{X,Y}=\sum_{i\in[0,N)}X_i\cdot Y_i$ to calculate the value of co-relations between consumers X and Y, where X_i or Y_i represents the value of X's or Y's ith element (i.e., the normalized value of consumer X's order amount for product i). The enhanced Pearson stresses the co-relations of consumers X and Y over products, that is, the co-relations will strengthen the similarity between the two consumers. And we further employ the Sigmoid function as $1/(1+\mathrm{e}^{-\frac{|U_{X,Y}|}{2}})$ to normalize $U_{X,Y}$.

Consider a set of m consumers C and a set of n products P. Based on (3), we can generate three input similarity graphs we need:

- Consumer order similarity graph (SimCO). Where $\mathrm{SimCO}_{u,v}=\mathrm{sim}(O_u,O_v)$, O_u or O_v is a vector donated by the amounts of orders placed by consumer u or v on each product in P;
- Consumer return similarity graph (SimCR). Where $\mathrm{SimCR}_{u,v}=\mathrm{sim}(R_u,R_v)$, R_u or R_v is a vector donated by the amounts of returns placed by consumer u or v on each product in P;
- Product return similarity graph (SimPR). Where $\mathrm{SimPR}_{u,v}=\mathrm{sim}(R_i,R_j)$, R_i or R_j is a vector donated by the amounts of returns placing on product i or j by each consumer in C.

In the proposed TARW, two trust relations and two trust networks will be formed based on input graphs (Intuition 1):

- Consumer return-trust relations $\omega_{u,v}$. The probability that consumer u trusts consumer v among all the consumers based on consumer return tendencies.
- Consumer latent return-trust relations $\alpha_{u,v,k}$. The probability that consumer u's return products trust consumer v's return products at step k based on product return probabilities.
- Consumer initial trust network $t_{u,v,0}$. The probability that consumer u trusts the consumer v based on consumer order preferences;
- Consumer trust network $\rho_{u,v}$. The probability that consumer u trusts the consumer v based on return patterns.

Intuitively, we propose to develop ρ based on the initial network t with the two trust relations α and ω.

4.2 Trust relations formation

First, the formula of the consumer trust relations based on return tendencies ω is given by

$$\omega_{u,v}=\frac{\mathrm{SimCR}_{u,v}}{\sum_{v\in[0,m)}\mathrm{SimCR}_{u,v}},\tag{4}$$

where SimCR is the consumer return similarity graph.

Then, we generate consumer latent trust relations α based on product return probabilities, where $\alpha_{u,v,k}$ can also be regarded as the stopping probability that consumer u stops at consumer v at step k in the process of the random walk. The formula for generating $\alpha_{u,v,k}$ is

$$\alpha_{u,v,k}=\sum_{i\in\mathrm{RP}_u}\sum_{j\in\mathrm{RP}_v}\frac{\mathrm{SimPR}_{i,j}}{\sum_{j\in[0,n)}\mathrm{SimPR}_{i,j}}\cdot\frac{R_{u,i}\cdot R_{v,j}}{(1+\mathrm{e}^{-\frac{k}{2}})},\tag{5}$$

where SimPR is the product return similarity graph, $\mathrm{SimPR}_{i,j}$ represents the similarity between the products i and j based on return probabilities. Besides, $R_{u,i}$ is the amount of returns between consumer u and product i, and $R_{v,j}$ is the amount of returns between consumer v and product j. And RP_u or RP_v is the set of return-products of consumer u or v.

Substantially, $\alpha_{u,v,k}$ is the trust-relation between consumer u's return products and consumer v's return products at step k. The parameter k represents the depth of the walk, when the depth of the walk increases, more noise and interference will occur in the process, and thus the stopping probability α will increase. Comparing with the TrustWallker model [25], the proposed TWAR introduces $\alpha_{u,v,k}$ to measure the similarity based on the union instead of the intersection of consumer u's and consumer v's return-product sets. In practice, products returned by consumer u and consumer v may be not the same but similar with each other according to the product return probabilities. With the stopping probability α, we can exploit the latent relevance between consumers u and v.

4.3 Trust network initialization

Prior studies on the trust network are mostly based on the Epinions' data sets. The Epinions allows consumers to mark other consumers as "trust" and "distrust" to obtain a consumer social trust network. However, most of the online retail platforms do not allow consumers to set up trust consumers, thus consumers' social network information is difficult to obtain in the real world.

The alternative to the social trust network must be found to apply this idea to the real world data. Azzedin et al. [31] defined the trust in the Internet as a firm belief held by an individual that what another individual will do. According to Figure 1, consumers can return products only after ordering. In other words, consumer's return-similar consumers can be addressed and identified in his or her order-similar consumers by collaborative filtering with the return similarity information.

Along these lines, the consumer order similarity graph is utilized and transformed to be the initial trust network $t_{u,v,0}$, where $t_{u,v,0}$ represents the probability that consumer u trusts the consumer v at step 0 based on order preferences. The parameter $t_{u,v,0}$ is given by

$$t_{u,v,0} = \frac{\mathrm{SimCO}_{u,v}}{\sum_{v\in[0,m)} \mathrm{SimCO}_{u,v}}. \tag{6}$$

4.4 Trust network formation

In the proposed model, what we finally obtain is ρ, which represents the consumer trust network over all the products based on return patterns. Here, $\rho_{u,v}$ represents the probability that the consumer u trusts consumer v based on return patterns, which is generated by the iteration of t in the random walk. For consumer u at the step k ($k > 0$), the iterative formula of $t_{u,v,k}$ is

$$t_{u,v,k} = \alpha_{u,v,k} \cdot t_{u,v,k-1} + (1 - \alpha_{u,v,k}) \cdot \omega_{u,v}. \tag{7}$$

Intuitively, $\alpha_{u,v,k} \cdot t_{u,v,k}$ in (7) is to further extend the latent return trust relevances from consumer u to v based on order preferences information. Besides, $(1 - \alpha_{u,v,k}) \cdot \omega_{u,v}$ adjusts the next iteration's trust relation from consumer u to v in t, based on the consumer return similarity in SimCR.

At step k, we can keep $t_{u,k}$, which represents consumer u's trust relations to other consumers. The iterative formula (7) will be terminated if $t_{u,k}$ is convergent, that is $|t_{u,k} - t_{u,k-1}| < 0.01$ in this paper. When the iteration is terminated at step k', we can get $t_u = t_{u,k'}$. Besides, according to the "six-degree of segmentation theory" [32], the maximum depth of one round walk is set to 6 steps in case that the iteration of t_u will never end. Finally, we obtain ρ: a m-by-m consumer trust network based on return patterns, by substituting all the consumers into the iteration of t_u and then regularizing t with $l1$ norm:

$$\rho = \frac{t}{\|t\|_1}. \tag{8}$$

5 TARW for business application

In this section, we present the two applications of the consumer trust network ρ: (i) return propensity estimation and (ii) consumer anomaly scoring.

5.1 Return propensity estimation

In this part, we try to estimate the return propensity $\hat{r}_{u,i}$ for consumer u on product i based on the consumer trust network ρ (Intuition 3), as we can see in Figure 4. Targeting at consumer u, we can obtain a top-l list of trust consumers sorted in a descending order according to the trust values in ρ_u, that is, we can get an l-size trust consumer set TC_u for consumer u. For consumer v in this trust consumer set, the trust relation from consumer u to v based on return patterns is $\rho_{u,v}$, and the amount of returns on product i is $r_{v,i}$. Hence, for product i, the estimation of consumer u's return propensity $\hat{r}_{u,i}$ is given by

$$\hat{r}_{u,i} = \sum_{v \in \mathrm{TC}_u} r_{v,i} \cdot \rho_{u,v}. \tag{9}$$

5.2 Consumer anomaly scoring

In this part, as we can see in Figure 4, we adopt the idea of anomaly detection in the RWR [24] to generate anomaly scores and categorize consumers (Intuitions 2 and 4), based on the consumer trust network ρ. Since ρ is one normalized graph, relations in this network are imbalanced as shown in Figure 5 (Intuition 1). In (10), we transform the directed edges in ρ into undirected edges by adopting the mean value of trust scores $\rho_{u,v}$ and $\rho_{v,u}$. The formula of the mutual-trust between the consumers u and v is shown below, where $\rho_{u,v}$ is an element in the consumer trust network ρ:

$$\text{mutual\_trust}_{u,v} = \frac{(\rho_{u,v} + \rho_{v,u})}{2}. \tag{10}$$

Based on the mutual-trust, we can identify the group types of consumers with their unique network-related properties. For example, given one product type, selfish consumers and honest consumers may appear similar order preferences, and we can distinguish the two groups with their return tendencies. That is, selfish consumers are likely to return more products than honest consumers (Nselfish > Nhonest), the return tendencies of selfish consumers are always stronger than those of honest consumers. Therefore, the mutual-trust between two selfish consumers is stronger than that between one selfish consumer and one honest consumer (high trust score > $\frac{1}{2}$ (medium trust score + high trust score)). And the latter is always stronger than that between two honest consumers (medium trust score). Likewise, fraud consumers are abnormal consumers who are different from others with respect to order preferences and return tendencies to all the products, and always account for the least among the return-consumers (Nfraud < Nhonest). In conclusion, given one product, the selfish consumers score the highest in the mutual-trust network, the second is the honest consumers, and then the fraud consumers.

For consumer u, the formula of anomaly score $\mathrm{as}_{u,i}$ on the product i is as follows, and here RC_i is the return-consumers of product i.

$$\mathrm{as}_{u,i} = \sum_{v \in \mathrm{RC}_i} \text{mutual\_trust}_{u,v}. \tag{11}$$

Apparently, returns placed by selfish consumers score the highest among the three groups. And returns placed by fraud consumers score the lowest among the three groups. Intuitively, consumers can be viewed as abnormal if their anomaly scores are considerably lower or higher than the average level among all the consumers. That is, the honest consumers are the normal consumers, while selfish and fraud consumers are abnormal. Based on the anomaly scoring formula, we can identify the return patterns of consumers for a product type.

6 Experimental results

In this section, four experiments are conducted to evaluate the performances of the proposed trust-aware random walk model (TARW). Here, experiments 1–3 evaluate the model's performances on the return propensity estimation, and experiment 4 conducts the study on the anomaly scoring. Besides, we initially chose 0.05 as the threshold, that is, we treated the predicted return propensity $\hat{r} \geqslant 0.05$ as "high return propensity" and $\hat{r} < 0.05$ as "low return propensity".

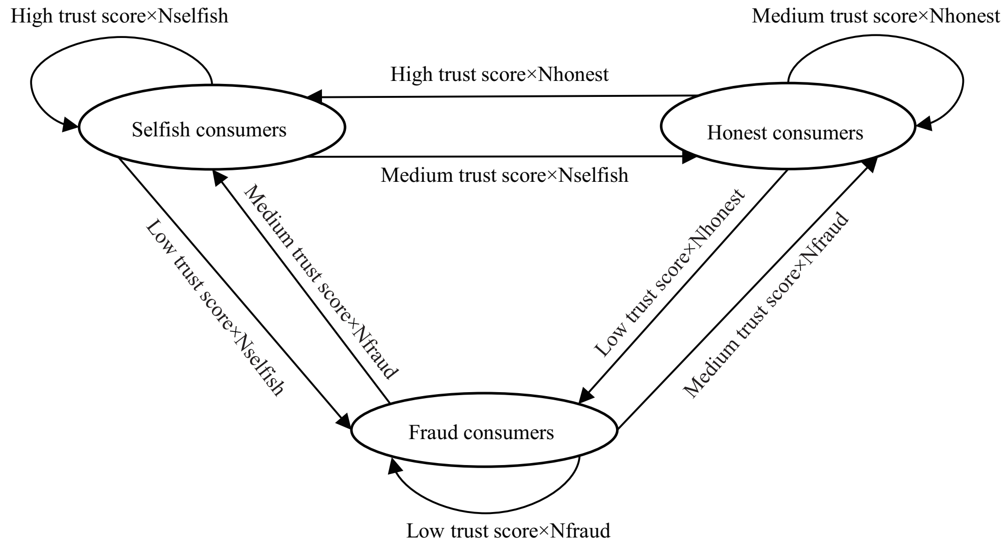

Figure 5 The imbalanced trust relations.

6.1 Evaluation on return propensity prediction

6.1.1 *Metrics*

Four metrics will be used in experiments 1–3: (i) MAE, (ii) Precision, (iii) Recall, and (iv) F1-score.

• **MAE.** MAE value (mean absolute deviation) is a metric to evaluate the accuracy of the predicted results. The smaller the value is, the higher the accuracy of the model's estimation tends to be. The formula of MAE is as follows:

$$\mathrm{MAE} = \sum_{u,i} |r_{ui} - \hat{r_{ui}}|/N, \tag{12}$$

where N is the test data set size, r_{ui} and $\hat{r_{ui}}$ are the observed value and the predicted value of consumer u's return amount on the product i, respectively.

• **Precision and Recall.** In practice, retailers always care more about the efficiency in finding out the return-potential consumers rather than the accuracy of the predicted results. And we measure the efficiency with the Precision and the Recall. Specifically, the number of the predicted return-consumers is set as N, the number of the actual return-consumers is set as n. The formulas of the Precision and the Recall are shown as

$$\mathrm{Precision} = \frac{|E_N \bigcap E_n|}{N}, \tag{13}$$

$$\mathrm{Recall} = \frac{|E_N \bigcap E_n|}{n}, \tag{14}$$

where E_N is the collection of the N-size predicted return-consumers, and E_n is the collection of the n-size actual return-consumers. $|E_N \bigcap E_n|$ is the number of consumers in both E_N and E_n.

• **F1-score.** F1-score is the harmonic average of the Precision and the Recall, which is used to evaluate the overall performance of the algorithm. The formula of F1-score is shown as follows:

$$\mathrm{F1} = \frac{2 \times \mathrm{Prec} \times \mathrm{Recall}}{\mathrm{Prec} + \mathrm{Recall}}, \tag{15}$$

where Prec is the Precision value of the algorithm, and Recall is the Recall value of the algorithm.

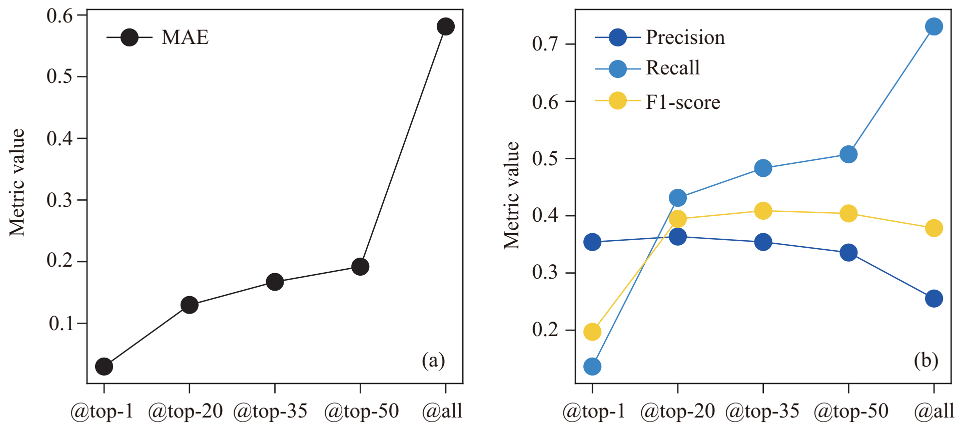

Figure 6 (Color online) Performance of different trust ranges. (a) MAE metric; (b) other metrics.

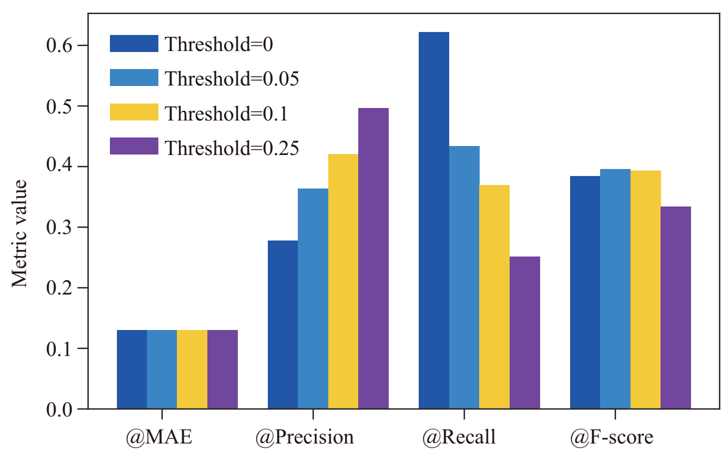

Figure 7 (Color online) Performance of different thresholds.

6.1.2 *Evaluations on trust range*

First, given the same threshold (0.05), we utilized the proposed model to estimate consumer return propensities with respect to five trust ranges (top-1, top-20, top-35, top-50 and all), and then we compared their performances. As shown in Figure 6, when larger trust range is adopted, the proposed model will be weaker at the Precision and the MAE, but the coverage reflected by the Recall will be greatly improved. When the trust range is 35, the F1-score measuring the comprehensive performance of the model is the highest, that is, the performances on the Recall and the Precision are well balanced. Intuitively, the top-35 of trust consumers are more likely to provide both creditable and diversified collaborative-information for retailers.

6.1.3 *Evaluations on threshold*

Second, we adopted the same trust range (top-20) and compared performances based on four different thresholds (0, 0.05, 0.1, and 0.25), as diagrammed in Figure 7. Apparently, the Precision is in positive correlation with the threshold, while the Recall is in negative correlation. When the threshold is 0.05, the F1-score is the highest. That is, when both efficiency and coverage are taken into account, if more than one (i.e., trust range $\times$ threshold: 20×0.05) returns in total were placed by the target consumer's top-20 trust consumers on a product, this consumer can be labeled as "return-potential" by retailers.

6.1.4 *Evaluations on model difference*

Third, we compared the performances of the proposed model (TARW) and the random walk with restart model (RWR), as shown in Figure 8. The trust range of the TARW was set as 20, the threshold was

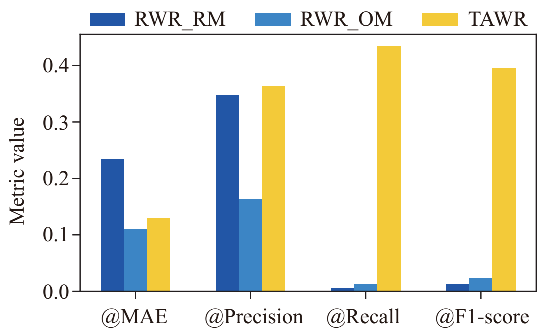

Figure 8 (Color online) Performance of different models.

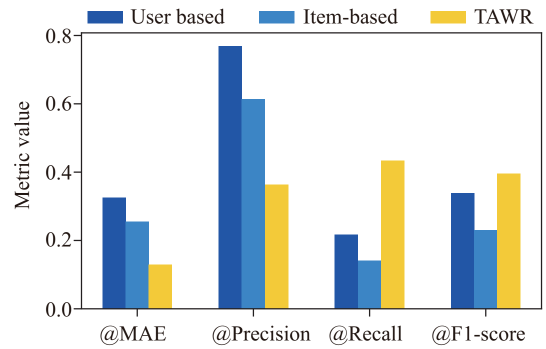

Figure 9 (Color online) Performance of baseline models.

Table 5 Description of purchase records

Product ID	Product price	Sales volume	Number of returns	Number of consumers
14064167845	88	377	53	318

set as 0.05, and the restart probability of the RWR was set as 0.15. We first applied the RWR on the return records (RWR_RM) and found limited trust consumers for each consumer. That is, RWR_RM is extremely weak at the Recall value, though the value of Precision is considerably high. Hence, the basic RWR model is weak at dealing with the data sparseness problem in practice. The RWR_OM is constructed by applying the RWR on the order records, and this model has better performance at the Recall but weaker performance at the Precision. The TARW outperforms both the RWR_OM and the RWR_RM at almost of the metrics. Therefore, it is feasible and effective for the TARW model to extract return-similar consumers from the order-similar consumers.

Finally, we compared performances of the proposed model and two baseline similarity models in Figure 9. Based on the same trust range and threshold, it is obvious that the TARW outperforms the item-based model (IF) and user-based model (UF) on the Recall at relatively little cost of the Precision, which ultimately leads to the better performance at the F1-score. To sum up, the proposed model outperforms other baseline models, and the ideas of the trust network and the collaborative filtering in the proposed model have greatly improved the efficiency of the RWR.

6.2 Study on anomaly scoring

In this section, we observed the anomaly scores to identify consumers' return patterns and segment consumers for one given product. As shown in Table 5, the given product's total sales volume was 493 during January 12, 2013 to January 12, 2014, which can be seen as popular among the products at this price level. Among the ordered-items, 70 items were returned or refunded. Figure 10 presents the distribution of order-consumers' anomaly scores, and then we extract the return-consumers from the order-consumers in Figure 10, where we set two baselines:

- **Order_baseline:** the (average − variance) value of anomaly scores among order-consumers for the given product, which is valued as $0.99 - 0.3 = 0.69$.
- **Return_baseline:** the (average + variance) value of anomaly scores among return-consumers for the given product, which is valued as $1.78 + 0.34 = 2.12$.

In this way, we can distinguish abnormal consumers who scores too low or too high among all the consumers. It is obvious that the average anomaly score of order-consumers for the given product is higher than that of return-consumers.

According to Figure 10, consumers are prone to cluster as three potential groups divided by the two baselines roughly. Specifically, consumers below the order_baseline are the fraud consumers; consumers between the order_baseline and the return_baseline are the honest consumers; and consumers above the return_baseline are the selfish consumers. With the horizontal contrast in Figure 10, we can see that few of the honest and fraud consumers choose to return the products while most of selfish consumers choose

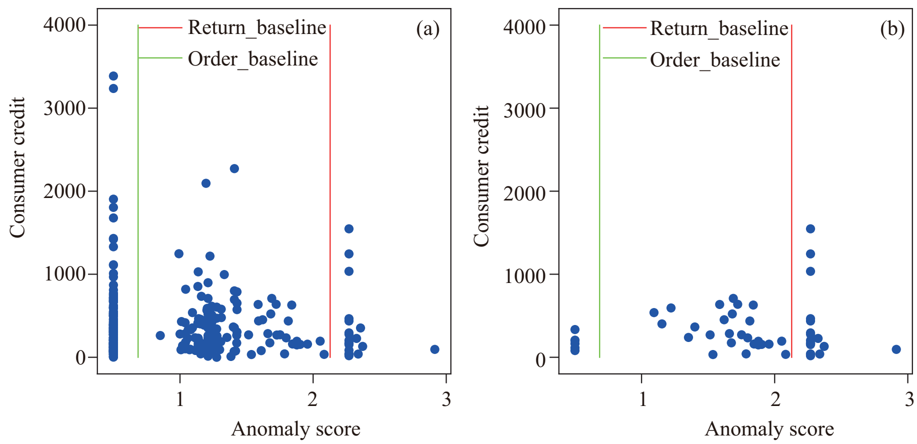

Figure 10 (Color online) Anomaly scores of order-consumers. (a) Among order-consumers; (b) among return-consumers.

Table 6 Description of consumer credit

Number of returns	Consumer's average credit	Percent (%)
0	400.5428	83.29
1	316.2121	16.22
2	116.5	0.49
Total	385.4717	100.00

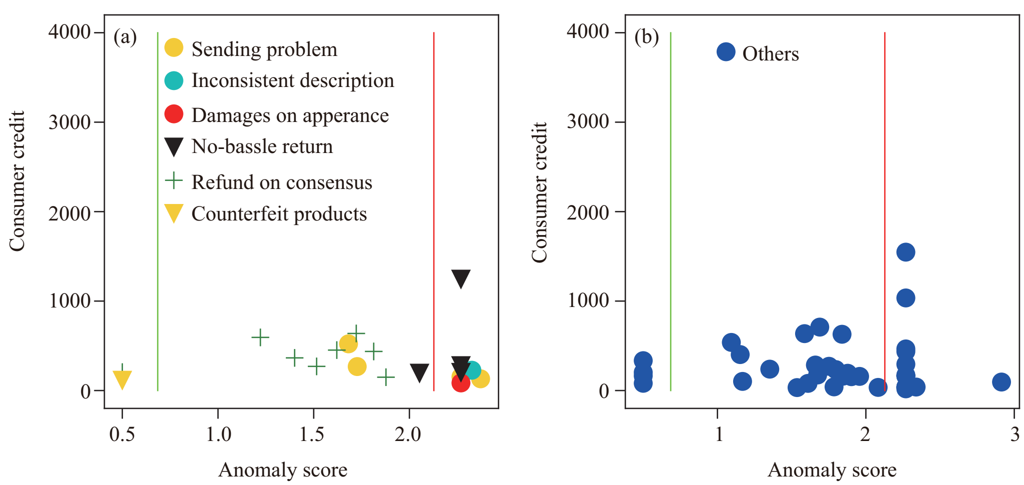

Figure 11 (Color online) Distribution of return reasons. (a) Specific return reasons; (b) unspecific return reasons.

to return. In fact, 39% of the returns are placed by selfish consumers who only account for 7% of all the consumers, and the return rate of selfish consumers is up to 90%. In addition, consumers of lower anomaly scores or higher credit scores appear weaker return tendencies, which is also shown in Table 6.

Further, we observed consumers' return reasons to provide authentic evidence about the proposed model's efficiency. And we can find in Figure 11, the reason "others" can hardly tell the groups of consumers. Besides, returns of selfish consumers are mostly motivated by avoidable factors that can be concluded as the expectation drop rather than quality issues. These returns can be predicted with the proposed model due to their high return propensities, which means they can be taken actions on before their returning. Moreover, returns placed by honest and fraud consumers are mostly due to retailers' mistakes such as quality problems and sending problems ("refund on consensus" refers to the situation that retailers and consumers reach the consensus on the request of refund). Returns placed by such consumers is difficult for the retailers to predict and control the loss. As for these consumers, retailers can check the authenticity of their return reasons and then carry out according measures.

Table 7 Comparison of credit

	Return-consumer credit	Order-consumer credit
Selfish-potential	338.9	328
Honest-potential	327.42	341.08
Fraud-potential	183.83	406.7

Table 8 Comparison of fraud consumers

Customer ID	Trade time (day)	Buyer credit	Active time (day)	Discount fee	Return num
l***9	6.944	81	613.53	0	3
D***L	0.059	164	548	31	1

Specifically, we conducted one extra study on the fraud consumers. With the horizontal contrast of return-consumers' and order-consumers' credit scores in Table 7, we can find that the average credit scores of selfish and honest consumers almost remain the same. And for fraud consumers, the average credit score of return-consumers is evidently lower than that of order-consumers, while the average credit score of the latter is the highest among the three consumer groups. As shown in Table 6, the average consumer credit is in negative correlation with the number of returns. Thus, we can infer that the detected fraud consumers are more likely caused by the ramp-up problem (i.e., returns related to those consumers are not sufficient enough to model their return patterns). However, some detected fraud consumers in return-consumers score relatively low in both anomaly scores and consumer credits, and these consumers may contain malicious users who post fake comments about the delivery or quality of the product. For instance, as we can see in Table 8, consumer D***L and consumer l***9 are both detected as fraud return consumers who selected "others" as the return reason. Consumer D***L is more likely to be a new consumer withdrawing the order right after attracted by discounts, who has consumed a lot but registered recently. Consumer l***9 may be the actual fraud consumer scoring low in "buyer credit", who has returned the product after actually obtaining it (according to "trade time").

To sum up, we can find that the return reasons of these three groups are roughly in line with the three return forces we noted before: (i) expectation gaps, (ii) retailer issues (e.g., product quality problem, inventory problem, etc.), and (iii) return frauds. It is feasible to segment consumers into several latent groups in terms of their return patterns (research question 1). With the auxiliary verifications in this section, we can find that the return pattern of different consumer groups will impact the return decision process of products (research question 2). For instance, most of selfish consumers choose to return products for avoidable reasons, thus it is feasible and effective to induce them to retain products. Returns placed by fraud consumers are mostly inevitable and harmful, retailers can block them from the source by marking them in the blacklists. By contrast, returns placed by honest consumers are inevitable but helpful, that is, retailers can observe honest consumers' feedbacks to perfect the products and service. Besides, it is effective for the proposed model to quantify consumers' return patterns with consumer order preferences and return tendencies (research question 3). The consumer categorization with the proposed model are helpful to understand the nature and underlying forces of consumers' return events. And with the proposed model, we can help retailers improve the conversion rates of selfish consumers, retain honest consumers, and block fraud consumers.

7 Conclusion

In this paper, we gained insights into the consumer return-force patterns by segmenting consumers as (i) selfish consumersa, (ii) honest consumers, (iii) fraud consumers. Then, we developed the TARW to capture the latent return patterns by generating a consumer trust network. To be specific, we used the consumer similarity graph based on order preferences as the alternative to the initial consumer trust network which is difficult to obtain in practice, so that we can apply the idea of the trust network on the real-world data and model consumers' characteristics in the order stage. After that, we demonstrated the feasibility of this change in experiments. Besides, we introduced the product similarity based on return probabilities as the stopping probability to exploit the latent relevances among consumers. With the proposed model, we can estimate the return propensities, identify the return patterns and segment consumers.

In practice, the proposed model can help retailers take appropriate actions to control the loss of returns. It is also demonstrated in experiments that the proposed model outperforms the RWR and other baseline models at most of the metrics.

Acknowledgements This work was supported by National Key R&D Program of China (Grant No. 2018YFB-1004300), National Natural Science Foundation of China (Grant Nos. 61773199, 71732002), and Philosophy and Social Science Foundation of Higher Education Institutions of Jiangsu Province, China (Grant No. 2017SJB0006).

References

1 Fu Y, Liu G, Papadimitriou S, et al. Fused latent models for assessing product return propensity in online commerce. Decis Support Syst, 2016, 91: 77–88
2 Dutta S, Biswas A, Grewal D. Regret from postpurchase discovery of lower market prices: do price refunds help? J Marketing, 2011, 75: 124–138
3 Potdar A, Rogers J. Reason-code based model to forecast product returns. Foresight, 2012, 14: 105–120
4 Wood S L. Remote purchase environments: the influence of return policy leniency on two-stage decision processes. J Marketing Res, 2001, 38: 157–169
5 Ang L, Dubelaar C, Lee B C. To trust or not to trust? A model of internet trust from the customer's point of view. In: Proceedings of BLED 2001, Bled Austria, 2001. 43
6 Zacharia G, Moukas A, Maes P. Collaborative reputation mechanisms for electronic marketplaces. Decis Support Syst, 2000, 29: 371–388
7 Dellarocas C. The digitization of word of mouth: promise and challenges of online feedback mechanisms. Manage Sci, 2003, 49: 1407–1424
8 Zhang Y, Liu T, Li R, et al. Evaluation model of buyers' dynamic reputation in e-commerce. Int J Multimed Ubiquit Eng, 2015, 10: 53–64
9 Grace A M, Williams S O. Comparative analysis of neural network and fuzzy logic techniques in credit risk evaluation. Int J Intell Inf Technol, 2016, 12: 47–62
10 Gunn S R. Support vector machines for classification and regression. ISIS Technical Report. 1998
11 Lee T S, Chiu C C, Lu C J, et al. Credit scoring using the hybrid neural discriminant technique. Expert Syst Appl, 2002, 23: 245–254
12 Jin L P, Dong J. Classification of normal and abnormal ECG records using lead convolutional neural network and rule inference. Sci China Technol Sci, 2017, 60: 078103
13 Jiang J, Li Y J, Feng Q Y, et al. A multiple user sharing behaviors based approach for fake file detection in P2P environments. Sci China Inf Sci, 2010, 53: 2169–2184
14 Massa P, Avesani P. Trust-aware collaborative filtering for recommender systems. In: Proceedings of International Conference on Cooperative Information Systems. Berlin: Springer Press, 2004. 492–508
15 Sarwar B, Karypis G, Konstan J, et al. Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on World Wide Web. Raleigh: ACM Press, 2001. 285–295
16 Shepitsen A, Gemmell J, Mobasher B, et al. Personalized recommendation in social tagging systems using hierarchical clustering. In: Proceedings of the 2008 ACM Conference on Recommender Systems. Lausanne: ACM Press, 2008. 259–266
17 Yang X W, Guo Y, Liu Y. Bayesian-inference-based recommendation in online social networks. IEEE Trans Parall Distrib Syst, 2013, 24: 642–651
18 Qian Y, Peng Z Y, Liang H, et al. A latent topic based collaborative filtering recommendation algorithm for web communities. In: Proceedings of 2012 Web Information Systems and Applications Conference. Hainan: IEEE Press, 2012. 241–246
19 Chen D E, Ying Y L. A collaborative filtering recommendation algorithm based on bipartite graph. Adv Mater Res, 2013, 756: 3865–3868
20 Rong H G, Zhou X, Yang C, et al. The rich and the poor: a Markov decision process approach to optimizing taxi driver revenue efficiency. In: Proceedings of the 25th ACM International Conference on Information and Knowledge Management (CIKM). Indianapolis: ACM Press, 2016. 2329–2334
21 Yu W. Analysis on trust influencing factors and trust model from multiple perspectives of online Auction. Cent Eur J Phys, 2017, 15: 613–619
22 Fouss F, Pirotte A, Renders J, et al. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. IEEE Trans Knowl Data Eng, 2007, 19: 355–369
23 Haveliwala T H. Topic-sensitive pagerank: a context-sensitive ranking algorithm for web search. IEEE Trans Knowl Data Eng, 2003, 15: 784–796
24 Sun J, Qu H, Chakrabarti D, et al. Neighborhood formation and anomaly detection in bipartite graphs. In: Proceedings of the 5th IEEE International Conference on Data Mining. Houston: IEEE Press, 2005. 8
25 Jamali M, Ester M. Trustwalker: a random walk model for combining trust-based and item-based recommendation.

In: Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Paris: ACM Press, 2009. 397–406
26 Jeh G, Widom J. SimRank: a measure of structural-context similarity. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Edmonton: ACM Press, 2002. 538–543
27 Zhang Z, Zeng D D, Abbasi A, et al. A random walk model for item recommendation in social tagging systems. ACM Trans Manage Inf Syst, 2013, 4: 1–24
28 Alexandridis G, Siolas G, Stafylopatis A. Accuracy versus novelty and diversity in recommender systems: a nonuniform random walk approach. In: Recommendation and Search in Social Networks. Berlin: Springer Press, 2015. 41–57
29 Gong J B, Gao X X, Cheng H, et al. Integrating a weighted-average method into the random walk framework to generate individual friend recommendations. Sci China Inf Sci, 2017, 60: 110104
30 Smith W R. Product differentiation and market segmentation as alternative marketing strategies. J Marketing, 1956, 21: 3–8
31 Azzedin F, Maheswaran M. Evolving and managing trust in grid computing systems. In: Proceedings of IEEE Canadian Conference on Electrical and Computer Engineering. Winnipeg: IEEE Press, 2002. 1424–1429
32 Travers J, Milgram S. The small world problem. Psychol Today, 1967, 1: 61–67