Received December 7, 2018, accepted January 8, 2019, date of current version February 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2893601

Mining Users Trust From E-Commerce Reviews Based on Sentiment Similarity Analysis

SHAOZHONG ZHANG[1] AND HAIDONG ZHONG[2]

[1]College of Electronic and Computer Science, Zhejiang Wanli University, Ningbo 315100, China
[2]College of Logistics and E-commerce, Zhejiang Wanli University, Ningbo 315100, China

Corresponding author: Shaozhong Zhang (dlut_z88@163.com)

This work was supported in part by the Natural Science Foundation of Zhejiang under Grant LY16G020012, in part by the Major Research Projects of Humanities and Social Sciences of the Ministry of Education of Zhejiang under Grant 2014GH015, in part by the Humanities and Social Sciences of the Ministry of Education under Grant 14YJC630210, in part by the Zhijiang Youth Action Project under Grant G306, and in part by the Open Research Funding Program of the Key Laboratory of Geographic Information Science of East China Normal University under Grant KLGIS2014A0l.

ABSTRACT Consumers' reviews in E-commerce systems are usually treated as the important resources that reflect user's experience, feelings, and willingness to purchase items. All this information may involve consumers' views on things that can express interest, sentiments, and opinions. Many kinds of research have shown that people are more likely to trust each other with the same attitude toward similar things. In this paper, we consider seeking and accepting sentiments and suggestions in E-commerce systems somewhat implies a form of trust between consumers during shopping. Following this view of point, an E-commerce system reviews mining oriented sentiment similarity analysis approach is put forward to exploring users' similarity and their trust. We divide the trust into two categories, namely direct trust, and propagation of trust, which represents a trust relationship between two individuals. The direct trust degree is obtained from sentiment similarity, and we present an entity-sentiment word pair mining method for similarity feature extraction. The propagation of trust is calculated according to the transitivity feature. Using the proposed trust representation model, we use the shortest path to describe the tightness of trust and put forward an improved shortest path algorithm to figure out the propagation trust relationship between users. A large-scale E-commerce website reviews dataset is collected to examine the accuracy of the algorithms and feasibility of the models. The experimental results indicate that the sentiment similarity analysis can be an efficient method to find trust between users in E-commerce systems.

INDEX TERMS E-commerce reviews, entity-sentiment word pair, sentiment similarity, trust.

I. INTRODUCTION

Reviews from consumers are very important information in E-commerce systems. Many online shops have developed reviews system for users to post their reviews. With the rapid development of social networking media, more and more people are willing to share their feelings, opinions and suggestions on their bought items with their friends or even strangers in social network applications or E-commerce systems. These reviews can be very useful for people's decision making in many different scenarios such as users' preference mining and personalized recommendation [1], [2]. At present, more and more review mining based applications are being applied to make our decision process easier than before. These applications have greatly changed people's behavior patterns, especially in E-commerce activities. For example, when people want to buy a product, book a hotel or restaurant, they usually not only ask for advice from their friends but also refer to reviews available online. To adapt to this change, many famous E-commerce companies, such as Amazon, eBay and Taobao(China), have built up well-function consumer review systems.

Online experience from various people can help one make decisions. In this case, people and their experience are required to be trusted by others. It makes sense that we usually ask for advice from our friends or family members before we make a decision. But the question is, why individuals are inclined to rely on strangers in cyber space to make decision? Scholars find a primary reason for that is their lack of trust in companies that they only experience through the web medium [3], [4]. The virtual nature of the web medium challenges traditional understanding of customer trust. In E-commerce scenario, customers have no chance to have a face-to-face interaction with a salesman or a direct physical experience with the store and the products they want to buy. On one hand, their experience is mediated through the web which is a two-dimensional graphical display. They usually feel somewhat lost and need someone to give them advices. On the other hand, reviews from consumers who purchase an item have direct physical experiences with it, are seem to be more reliable than vendor's promotions or advertising words.

However, E-commerce websites usually accumulate large scale text based reviews which records historical commentary about one subject or item. Usually, consumers are unable to distinguish which reviews can be trusted under so large information. Different consumers can hold different aspects and standpoints in viewing things. And their attitudes, interests, preferences, etc. will vary greatly towards the products or services. Some users give a positive rating because they like certain attributes of the product, while others give a negative rating because they don't like these attributes. Therefore, it is impossible for a consumer to judge whose reviews are suitable and which users can be trusted. The consumers urgently need to be established a trust between other users, which give the reviews he can trusts, provide him with an opinion reference, and shield the untrusted comments to prevent misleading to the user when he wants to purchase an item [1], [5].

Many scholars have spent much effort on the phenomenon of trust relationship between strangers in E-commerce environment and found an interesting result: people are more willing to trust the individuals who are similar with them in as many respects as possible. The similarity factors include the brought items, the sentiment style of reviews, the words used, etc. [6]–[8]. There are many studies trying to explore the relationship between people's mutual trust and their similarity quantitatively, and find that there is a strong correlation between both trust and interest similarity [9]–[12]. Although there is a certain relationship between the trust of users and the similarity of users, this relationship is not an obvious linear relationship between trust and similarity, and it also includes many other influential factors. How to correctly find the relationship between trust and similarity still faces great challenges.

Due to its human-related properties, trust is difficult to be uniformly defined or even precisely described. The vast majority of existing studies focused on trust construction and maintenance between customers and companies over time and after repeated experiences. While limited effort is spent on trust between consumers and potential consumers in E-commerce systems. Obviously, in the field of E-commerce reviews, people are more concerned about the credibility of reviews and the trust of user who post the reviews. In our work, we aim to investigate trust between users in E-commerce systems quantitatively by exploring their reviews and evaluations regarding to various commodities, services, businesses, and other related subjects. Based on sentiment analysis of large-scale text reviews in E-commerce website, we focus on sentiment similarity between users to establish their trust, which can provide potential support for further implementation of trust related recommendation service.

The rest of the paper is organized as follows. In section II, we described the related works of our study, including the existing studies in trust and its computation, issue of sentiment and similarity analysis, correlations between trust and similarity. In section III, we present some definitions and explanations related to sentiment similarity and trust. In section IV, we address the computation framework and process of users trust relationship based on sentiment similarity. In section V, a general approach of user direct trust computation is proposed based on sentiment similarity mining. Also, detailed steps of sentiment analysis, and users' propagation trust relation exploration algorithms are described. In section VI, we conducted the experimental evaluation to the proposed algorithms using public available E-commerce website reviews dataset. In the last section, we conclude the paper and discuss with possible points which need to be furthered in the future.

II. RELATED WORKS

A. TRUST COMPUTATION

The concept of trust has been studied by scholars all over the world in diverse contexts and from various disciplines, including Economics, Management, Computer Science and Sociology [8], [13], [14]. In the field of economic research, trust is explained from the theory of rational choice, and defined as a rational action to make a choice of whether to give a trust after careful thought and cost calculation. That is, individuals are inclined to make rational and maximizing benefit choices, which is usually called computational trust in existing research work [1].

In trust computation, there are two core kinds of trust as direct trust and indirect trust [15]. Trust can be established according observations on whether the previous interactions among the subjects, and can be called direct trust. Direct trust is used for reflecting the trustworthiness between direct connected users. Typically, many direct historical interactions data, such as behaviors, reviews, or other various evidences, are used to compute the direct trust degree among users [16], [17]. Dimah explore the potential of social information derived from micro-bloggings as a source of user relevant recommendations. They proposed an approach ISTS that can exploit two factors from online social network: the sentiment orientation in friends posts about certain items and the trust relations between friends [18]. Li and Dai [19] proposed a promising methodology to handle the trust mechanism for P2P network. They let parties rate each other after the completion of transaction, and use the aggregated ratings of a given party to derive a trust score.

While indirect trust is used widely in long path connected users through intermediate users [16]. This kind of trust is treated as transferability that may arise from one familiar context to another new context, or from one trusted entity to another unknown entity. Bo, Yang. and Qiang present an information propagation mechanism in semantic web to semantic trust score computation. Entire trust is measured by a combined trust score from both subjective and objective sides of information [20]. The objective side of trust is semantic trust of information, and the subjective side is trust relationship between peers. And trust relationship is based on the historical interactions between peers [21]–[23]. Faruk and Arnab proposed a trust management model which will take factors like direct, indirect and global trust of the service to find out the final trust value of the service. They take the trust path distance into consideration while calculating the indirect trust [24]. Li *et al.* [25] propose two trust models to content delivery network: a local trust model and a cross-domain trust model. Hong *et al.* [26] presents a new method called Max-aggregation, which calculates peer's reputation through the view of multidimensional and multi-attribute, and get the indirect reputation using trust propagation and aggregation in the web of trust.

B. SENTIMENT SIMILARITY ANALYSIS

The methods take the similarity analysis as an important and basic content, which consider the sentiment and emotion as the evaluation factors for trust. Additionally, sentiment and affective similarity analysis have been studied extensively in natural language understanding, data mining and statistical analysis [27]–[29].

The existing methods of sentiment analysis based similarity exploration can be divided into three levels, which are document level, sentence level, and entity and feature level. All three levels are based on opinion lexicon, which is a collection of specific keywords or sentiment lexicons (extracted from gathered reviews) with parts-of-speech tags and treaded as the basis for analyzing the reliability of reviews [30]. At document level, the task is to classify a whole opinion document according to whether it expresses a positive or negative sentiment [31], [32]. At sentence level, the task considers the sentences and determines whether each sentence expresses a positive, negative, or neutral opinion. Neutral usually means that no opinion is given [33], [34]. The analysis, both at the document level and at the sentence level, cannot exactly discover those specific objects whether people like or dislike. At entity and feature level, the approach concerns directly about the opinion itself instead of looking at language constructs (documents, paragraphs, sentences, clauses, or phrases). It is based on the idea that an opinion consists of a sentiment (positive or negative) [35]. Hsu [2] adopted a sentiment word database to extract sentiment-related data from microblog posts and used these data to investigate the effect of different types of sentiment-related words on product recommendations.

The main goal of the existing sentiment analysis methods is to cluster the sentiments of users, commonly dividing people's sentiments to things into several types. Even at the entity and feature levels, its main purpose is to divide the user's sentiments into likes or dislikes. However, the above methods concern directly in overall trend which is insufficient when we calculate the trust based on sentiment similarity. It is necessary to analyze the specific attitudes on specific objects in reviews.

C. ON CORRELATIONS BETWEEN TRUST AND SIMILARITY

Over the past few years, many works have focused on the relationship analysis between trust and similarity. The similarity analysis based on sentiment has become an important research approach to establish trust relationship. Many studies have shown that there is highly correlation between trust and similarity. They demonstrated that individuals with similarities also have a high degree of trust in certain areas. These similarities include interest, content, behavior, etc.

Cai-Nicolas Ziegler and J. Golbeck investigated correlations between trust and interest similarity. They established a formal framework for investigating interactions between trust and similarity. They used a mathematical model to compute similarity and presented computation algorithms for profile and profile similarity. They used two experiments to analyse possible positive correlations between similarity and interpersonal trust. At meanwhile, through the analysis of the data from the FilmTrust Web site, the results show that when the similarity of users changes within a certain range, the trust between users changes accordingly. This change indicates that there is very strong relationship between trust and similarity [10]. Li proposed a node interest similarity based trust model, which took both node interest bias and reputations in each interest domain into consideration, and used interest domain reputation vector to maintain the behaviors of node in specific interest domain. They used interest similarity between nodes to weight domain local trust recommendation [36]. These innovative studies proved that there is a correlation between trust and similarity, and they had presented the corresponding calculation method.

Other relevant important studies including: Zhang *et al.* [37] proposed a contextual trust evaluation method by comparing the transaction context similarity between the new transaction and past transactions. Gang *et al.* [38] proposed a recommendation trust model, which is based on E-commerce transaction content similarity and differentiates the trust degree of acquaintance node recommendation from stranger node recommendation. Golbeck studied the trust and nuanced profile similarity in online social networks. They investigate features of profile similarity and how those relate to the way users determine trust. Through a controlled study, they isolated several profile features beyond overall similarity that affect how much subjects trust a hypothetical user [39]. Hossein proposed a method that employed user similarities to extract trust values without any need of direct rating. User similarity is calculated from profile information and shared text via text-mining techniques [40]. Melika proposed a method that is tried users rating of certain areas to be gathered and the similarity of users or items are measured, and realized and recommended the most suitable and nearest item of user's preference. Then the available approach to measure similarity is recommended to the target user and the trusted user will be found [41]. Zhang *et al.* [42] presented a behavior similarity inspection module to trust inspection according to the behavior similarity rule of the human society and the definition of the mathematic probability. Guoming *et al.* [43] presented a conspiracy group detection trust model based on behavior similarity.

However, most of the existing studies on the relationship between similarity and trust are based on profile similarity computation, interest or transaction content, and there is no research on trust based on sentiment similarity.

D. PROBLEMS AND OUR SOLUTIONS

However, most of these above methods focus on exploring the overall trend of some sentiment or emotional tendencies to classify users by text sentiments. They do not take into account the similarity of sentiment between individuals and trust relationship between users. And at meanwhile, sentiments can spread in social networks, and this propagation feature also has an important impact on user relationships. So, if we want to get accurate text similarity and user trusts, we should start from real sentiment vocabulary [44], [45]. Bloom believe that in the reviews text, the sentiments words are always targeted at specific evaluation objects [46]. The extraction of sentiments words and their evaluation objects is the basis and important task in the research of sentiment mining. This is because the evaluation object sentiment word combinations contain more information than either of them alone.

Different from traditional trust modules and sentiment analysis, these are the following considerations in our study:

(1) for sentiment similarity computations, we use a deep and more granular division to the reviews text. However other traditional sentiment analysis studies were able to find the propensity of sentiments, but this tendency concern in the overall evaluation and trend of the review. These cannot reflect the perception of the specific attributes and characteristics of things in a reviews. We propose a fine-grained analysis method for the evaluation entity-sentiment word pairs by extracting the specific attribute words and feature in the reviews.

(2) for direct trust computation, that is, one to one trust in the work, we use the weighted average method to compute them, which is similar to other existing works. However, at the same time, we introduce an accompanying factor of sentiment, the rating which widely exists in E-commerce reviews, for weights evaluation. Which is, the direct trust calculation impacted by the facts whether the users have the same sentimental tendency or not for the same thing.

(3) for propagation trust computation, which is one to one trust though a third ones, we introduce graph based propagation algorithm. based on the proposed trust representation model, we use a shortest path to describe the tightness of trust and put forward an improved shortest path algorithm to figure out propagation trust relationship between users. The propagation trust is computed by integrating the direct trust based on shortest path algorithm.

III. SENTIMENT SIMILARITY AND TRUST RELATED DEFINITIONS

A. SYMBOLS AND DESCRIPTION

To facilitate description, we list all the nomenclatures involved in this section.

(1). U and $u_i \in U$, U is a set of user in E-commerce and u_i is one of them.

(2). O and $o_{u_i} \in O$, O is the set of objects that user give reviews on and o_{u_i} is an object set that of user u_i.

(3). T and $t_{u_i} \in T$, T is the reviews text set and t_{u_i} is the text that given by user u_i.

(4). E and $e_i \in E$, E is set of entity words in review text T about objects O.

(5). S and $s_i \in S$, S is set of sentiment words in review text T about objects O.

(6). $MI(e_i, s_j)$, indicates the degree of association between entity words and sentiment words.

(7). $p(e_i, s_j)$, is the joint probability of e_i and s_i.

(8). $p(e_i)$ and $p(s_j)$, $p(e_i)$ and $p(s_j)$ are marginal probability respectively.

(9). $r = (e, s) \in R$, R is score of an entity-sentiment words pairs mining from reviews text T.

(10). $TF_{i,j}$, is the frequency of the entity-sentiment word pairs e_i and s_j in text.

(11). IDF, is the frequency of the entity-sentiment word pairs in reviews text.

(12). $\theta(s_j)$, is influence factor that affect the vocabulary expressing degree of adjectives and adverbs.

(13). w and w', w is the degree of the entity-sentiment word pairs r, w' is the degree with influence factor to w.

(14). $\bar{r}_i$, is represents the mean score of entity-sentiment word pairs in review text of user u_i.

(15). $\vec{r}_i$, is represents the score vector of entity-sentiment word pairs in review text of user u_i.

(16). $Sim(t_{u_i}, t_{u_j})$, is the score of sentiment similarity between two reviews text given by u_i and u_j respectively.

(17). $\rho(i)$, is a control function of the number of reviews by users.

(18). s_{u_i}, is the rating score on an object of users u_i.

(19). $\sigma(s_{u_i}, s_{u_j})$, is an influence function from rating of u_i and u_j on object.

(20). D_{u_i,u_j}, is the direct trust degree of user u_i and u_j.

(21). N, is the number of paragraph in a review.

(22). n_k, is number of paragraph in which the word appears in a reviews.

(23). G, is trust network graph established by direct trust links and G' is a subgraph of G.

(24). PR_{u_a,u_b}, is propagation trust of two users where there is no direct trust.

B. DEFINITIONS

Definition 1: Sentiment similarity. There are two reviews text t_{u_i} and t_{u_j}, which are posted by user u_i and u_j on an object respectively. The sentiment similarity between t_{u_i} and t_{u_j} can be denoted as $Sim(t_{u_i}, t_{u_j})$.

The sentiment similarity of reviews of two different users is firstly based on extracting the entity-sentiment word pairs of each review, and then performing similarity computing on the entity-sentiment word pairs. Entity-sentiment pair is a combination of entity and sentiment word. Each review has its own entity-sentiment pairs, which one part is formed with object of evaluation that is a noun, a noun phrase or a clause, which acts as a subject, object, subject clause or object in the sentence, and the other one is sentiment word that express a certain emotional polarity. Most sentiment words are adjectives, adverbs, etc. From the previous analysis, If the two reviews text have similar entity-sentiment word pairs, then they can be deemed as similar. Therefore, sentiment similarity of the two reviews text can be measured by their entity-sentiment word pairs.

The correlation of trust and similarity is not surprising. A connection between user similarity and trust was established [10]. This work showed that there was a strong and significant correlation between trust and similarity. It showed that the more similar two people were, the greater the trust between them.

Relevant evidence from the study of social psychology shows that there is a strong relationship between trust and similarity. However, similarity cannot be the same as trust. This is because when people express their opinions and feelings, their evaluations of things vary widely. On the one hand, these reviews have the same meaning, such as widespread support or opposition to something. This commonality indicates the consistency of attitudes, sentiments, and emotions, and it can be obtained from the sentiment similarity. At the same time, the users' reviews words are also in various. These different vocabularies represent the users' different sentiments and perspectives, and this is where the users' reviews are different. Finding out these differences and determining the impact of consistency are the problems that should be considered in the direct trust relationship calculation.

Definition 2: Direct trust. If $o_{u_i} \cap o_{u_j} \neq \Phi$, there is a direct trust between user u_i and u_j, expressed as D_{u_i,u_j}.

Where $o_{u_i} \cap o_{u_j}$ is a set of objects that the two users have reviewed simultaneously. The two sets represent the set of objects that user u_i and user u_j review on respectively. If the two sets are not empty, that is, the two users have the same objects that are reviewed, the direct trust of the two users can be computed through the sentiment similarity of the reviews.

By analysing the two users' reviews on the same objects or services, the direct trust between them can be calculated. When there is such trust between two users, we believe that there is a direct trust link between the two users. u_i and u_j are treated as nodes of user, which represent any two different users in E-commerce system. We assume each of them posted some reviews on E-commerce objects. These objects include merchandise, merchant or service. These reviews may be aimed at one or more objects, and the users may have made multiple reviews. In these reviews, The users u_i and u_j may have some reviews on same objects. For example, user u_i bought a certain type of cell phone and gave a reviews and rating on it, at meanwhile the other user u_j is also like him. These reviews from u_i and u_j can be used to analyse the sentiment similarity between them.

We can use links to connect those users who have direct trust and build a trust network graph, denoted by G. However, usually in E-commerce system, not all users have common objects and give their reviews, that is when $o_{u_i} \cap o_{u_j} = \Phi$, there are still have a large number of them who have never given their reviews on the same objects. In this case, the trust relationship of these users cannot be calculated based on sentiment similarity directly and it is impossible to carry out direct trust links between them. The computing of sentiment similarity cannot be carry out and the direct trust will not be conducted. We can use the direct trust relationship between several users with direct trust links to the two users to calculate their trust. These several users are intermediate users who can propagate trust relationships within the two users who are no direct trust links.

Definition 3: Propagation trust. trust network graph G is established by direct trust. It consists of nodes and edges. The nodes represent users and edges represent direct trust link. There is a subgraph G' in G, where includes user nodes u_a and u_b. If there is at least one trust path composed of trust links in the subgraph, so that the two users can connected, it is said that there is a propagated trust relationship between the two users u_a and u_b, denoted as PR_{u_a,u_b}.

According to the transitive feature of trust, propagation trust can be obtained through the transfer relationship of other user nodes in a graph, as shown in figure. 1. The u_a and u_b are two users who are not direct related trust. They only have direct trust with some other users in the trust network. In this case, the two users will not sure having no trust relationship between them. They can build a propagate trust through direct trust of other users.

IV. SENTIMENT SIMILARITY BASED USER TRUST RELATIONSHIP CALCULATION FRAMEWORK

The users are usually the consumers who have involved in E-commerce activities. They may have purchased some items or services and posted reviews on these objects, as shown in figure 2. Typically, a user can post multiple reviews on multiple items. Therefore, these reviews for specific items can be expressed in several texts. These reviews can usually be obtained by collecting network information.

To find the trust, including direct and propagation, based on sentiment similarity of reviews by users in E-commerce

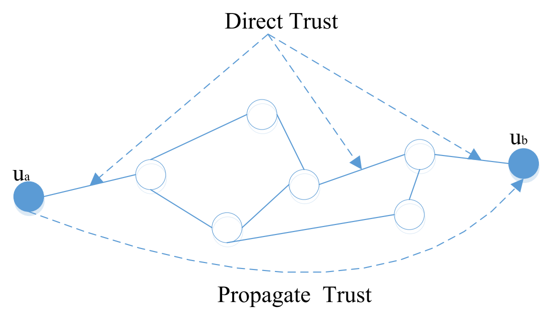


FIGURE 1. Propagation trust link based on direct trust.

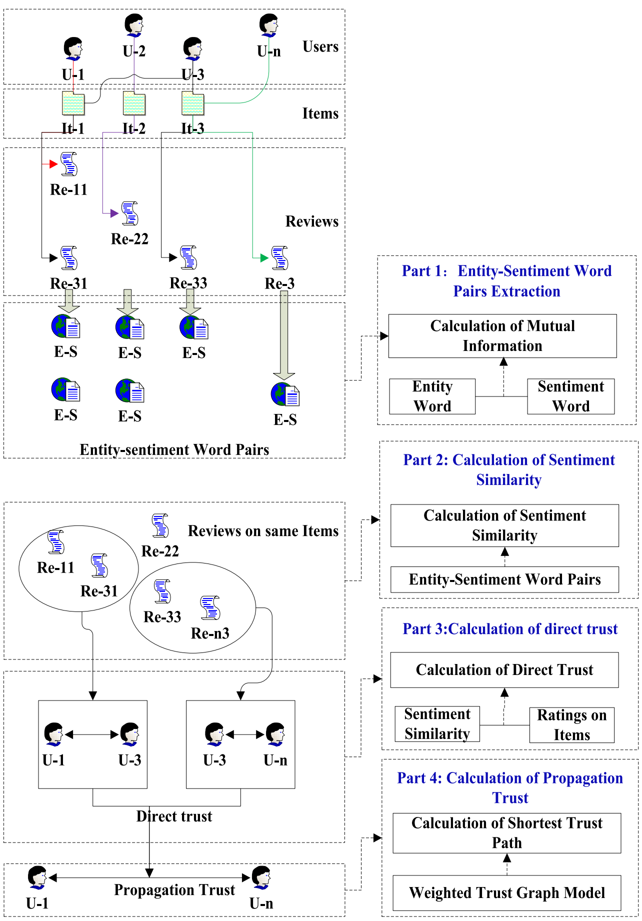


FIGURE 2. Trust calculation framework based on sentiment similarity.

systems, we propose a generally four-step computing framework.

Firstly, the entity-sentiment word pairs are extracted from reviews. The step is a key process to further deal with the sentiment similarity analysis and direct trust computing. The extraction of entity-sentiment word pairs in part 1 is mainly to analyse the vocabulary of the text, extract the entity words and sentiment words which describing the object. We use NLProcessor linguistic parser for entity word and high frequency word combined with public lexicon for sentiment word. A mutual information formula is used to calculate the relationship between each entity word and the sentiment word. And then we can find those words pairs with close relationship.

Secondly, we perform sentiment similarity calculations for two user-related reviews based on entity-sentiment word pairs. Calculation of sentiment similarity in part 2 is to compute the similarity degree of different reviews texts. This step is to use the obtained entity-sentiment word pairs for comparative analysis.

Thirdly, we use a new formula to calculate direct trust between two users with a common review objects. Calculation of direct trust in part 3 is to compute direct trust whom have reviews on same item or object. The calculation method mainly includes two aspects, one is sentiment similarity, and the other is the user's rating of the object.

Finally, calculation of propagation trust in part 4, assume users as nodes, and the links, which are based on their direct trust, as edges to create a trust network. Then we use an improved shortest path algorithm to calculate the propagation trust links between each pair of user nodes.

V. CALCULATION METHOD OF TRUST

A. ENTITY-SENTIMENT WORD PAIRS EXTRACTION

In entity-sentiment word pairs extraction, the entities usually are nouns or noun phrases which represent some specific objects, features, or attributes, etc. The sentiments are adjectives or adverbs which express emotions, opinions, or tendencies, etc. We apply the association rules to extract frequently occurring nouns or noun phrases as entities, and we use the adjectives or adverbs as sentiment which have the closest information distance to the object.

The extraction of entity word is achieved by part-of-speech tagging. In review sentences, the attribute of an entity is usually represented by a noun or a noun phrase. We used the NLProcessor linguistic parser to parse each review to split text into sentences and to produce the part-of-speech tag for each word. NLProcessor outputs linguistic information by directly marking text with XML tags: tokens are represented as "W" elements, word-class part-of-speech information is provided in their "C" attribute, noun and verb groups are marked as NounGroup and VerbGroup elements and sentences are marked with "S" elements. For example,

<S>
<NounGroup><W C=NNP>**Children**</W>
</NounGroup>
<VerbGroup><W C=VBZ>**love**</W>
</VerbGroup>
<NounGroup><W C=NNS>**toys**></W>
</NounGroup>
</S>

The extraction of sentiment word is achieved by high frequency words appearing in the reviews text. For each sentence in the review dataset, if it contains a frequent feature, we extract all the adjectives and adverbs. At the same time, we also consider using the existing lexicon of sentiment words, such as WordNet. If the word belongs to the lexicon, we extract them as sentiment word though it is not a high-frequency word in the reviews.

The extraction of entity-sentiment word pairs is implemented by the mutual information algorithm that being existed and commonly used in information process field. The mutual information method can be used to represent the relationship between two kinds of information or measure of the statistical relevance of two random variables.

Different word pairs in entity-sentiment word pairs are different from the sentiment strength expressed, and it is necessary to describe the degree of these word pairs. the degree of entity-sentiment word pairs can be measured by *TF-IDF* (Term Frequency–Inverse Document Frequency) method [47].

Mutual information between an entity word e_i and a sentiment word s_j can be calculated as formula 1.

$$MI(e_i, s_j) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(e_i, s_j) \log \frac{p(e_i, s_j)}{p(e_i) \cdot p(s_j)} \tag{1}$$

where $p(e_i)$ is the probability of occurrence of the entity word e_i, $p(s_i)$ is the probability of occurrence of a sentiment s_j and $p(e_i, s_j)$ is the joint probability that e_i and s_j occuring at the same time. If e_i and s_j are independent or irrelevant, $p(e_i, s_j) = p(e_i) \cdot p(s_j)$. The function $MI(e_i, s_j) = 0$ indicates that the entity e_i and sentiment s_j are irrelevant, and otherwise they are related. The greater the $MI(e_i, s_j)$, the closer the relationship between them. When the value of $MI(e_i, s_j)$ exceed a certain number, it indicates that there is a mutual association between two words. Then the two words can be regarded as an entity-sentiment word pair. That is $r = (e, s) \in R$, R is the set of entity-sentiment word pair.

The weight of entity-sentiment word pair is described by degree and is carried out with the *TF-IDF*, and it can be expressed as formula. 2

$$w_{i,j} = TF_{i,j} \times IDF = \frac{n_{i,j}}{\sum_{k=1}^{N} n_{k,j}} \times \log \frac{N}{n_k} \tag{2}$$

where $TF_{i,j}$ is the frequency of the entity-sentiment word pairs e_i and s_j in a single text, and IDF is the frequency of the entity-sentiment word pairs in all reviews; N is the number of paragraph in a review and n_k is number of paragraph in a review in which the word appears.

In many cases, vocabularies, such as adjectives and adverbs play a significant role in most of the reviews. Therefore, we set an influence factor to describe the vocabulary expressing degree of adjectives and adverbs to be analysed. The influence factor is denoted as $\theta(s_j)$, and can be expressed as formula 3.

$$\theta(s_j) = 1 + \log(1 + \frac{n_j}{N} \times \lambda(s_j)) \tag{3}$$

If $\theta(s_j)$ is a vocabulary that express degree, then $\lambda(es_{i,j}) = 1$, otherwise it will be 0. n_j and N have the same meaning as they are in formula (2).

According to formulas (2) and (3), w' can be acquired with the following equation and w' is a new degree with influence factor. w' is expressed as formula 4.

$$\begin{aligned} w'_{i,j} &= TF_{i,j} \times IDF \times \theta(s_j) \\ &= \frac{n_{i,j}}{\sum_{k=1}^{N} n_{k,j}} \times \log \frac{N}{n_k} \times \left[1 + \log\left(1 + \frac{n_j}{N} \times \lambda(s_j)\right)\right] \end{aligned} \tag{4}$$

Considering consumers usually make only one review on each item they bought, the paper treats N as the number of paragraphs of a text, which have independent sentiments. Meanwhile, n_k is the set to the same value as the number of sentences where predefined characteristic word appears. Therefore, the extracted entity-sentiment word pairs can be expressed as a vector, shown in formula 5.

$$\vec{r}_i = ((r_1, w'_1), (r_2, w'_2), \ldots (r_i, w'_i)) \tag{5}$$

B. SENTIMENT SIMILARITY AND DIRECT TRUST COMPUTING

The sentiment similarity is to compute the similar degree of two reviews texts. For two users u_i and u_j, the reviews texts they post on an object o are t_{u_i} and t_{u_j} respectively. The similarity degree between these two texts can be calculated by formula (6) based on Pearson Correlation.

$$Sim(t_{u_i}, t_{u_j}) = \frac{\sum_{k=0}^{\min|V|} (r_{i_k} - \bar{\bar{r}}_i)(r_{j_k} - \bar{\bar{r}}_j)}{\sqrt{\sum_{k=0}^{\min|V|} (r_{i_k} - \bar{\bar{r}}_i)^2 \cdot \sum_{k=0}^{\min|V|} (r_{j_k} - \bar{\bar{r}}_j)^2}} \tag{6}$$

where $\min|V|$ is the minimum number of entity-sentiment pairs in reviews text of user u_i and u_j. For any two users who have reviewed on same object, their sentiments may be positive, negative, or neutral. Whenever the emotion or sentiments of their reviews on the same object are consistent or approximate, the opinions or sentiments of the two users are treated as generally similar in E-commerce environment. Thus, sentiment similarity between users u_i and u_j can be acquired based on the context of their all reviews on same object, and can be calculated according to the following formula (7).

$$\begin{aligned} D_{u_i,u_j} &= \frac{1}{|O_{u_i} \cap O_{u_j}|} \\ &\quad \times \left(\sum_{i \in O_{u_i} \cap O_{u_j}} \rho(i) \frac{\sum_{k=0}^{|T|} \sigma(S_{u_i}, S_{u_j}) \cdot Sim(t_{u_i}, t_{u_j})}{\sum_{i \in O_{u_i} \cap O_{u_j}} Sim(t_{u_i}, t_{u_j})} \right. \\ &\quad \left. + (1 - \rho(i)) \cdot D_0 \right) \end{aligned} \tag{7}$$

where $O_{u_i} \cap O_{u_j}$ is a set of objects that the two users have commented simultaneously. D_{u_i,u_j} is the direct trust degree of user u_i and u_j. $|T|$ is the number of reviews, ρ and σ is the adjustment function.

In E-commerce, user ratings are ubiquitous information. We consider using user rating factors and combining sentiment similarity to calculate users trust. $\sigma(s_{u_i}, s_{u_j})$ is an influence function and it can be calculated using a covariance function as following formula 8.

$$\sigma(s_{u_i}, s_{u_j}) = \frac{\sum_{i=1}^{n}(s_{u_i} - \bar{s}_{u_i,u_j})(s_{u_j} - \bar{s}_{u_i,u_j})}{n-1}, \quad n \neq 1 \tag{8}$$

where s is the rating for an object. s_{u_i} and s_{u_j} indicate that the rating on an object of users u_i and u_j respectively, and $\bar{s}_{u_i,u_j}$ is the mean of the ratings of the two users on same object.

$\rho(i)$ is a control function of the number of reviews by the two users. It reflects the user's activity level and stability to the objects. The higher the control function, the greater the degree of intersection between the two users, the higher the activity level, and then the greater the influence of similarity on trust. Conversely, the smaller the control function, the smaller the degree of intersection between the two users, the lower the activity, and the less the effect of similarity on trust. The control function has the following properties:

(1). $\forall i$, $\rho(i) \in (0, 1]$

(2). $\rho(i)$ is monotonically increasing function, that is when $i_1 < i_2$, there is $\rho(i_1) \leq \rho(i_2)$;

(3). When $i \to \infty$, there is $\rho(i) \to 1$

The calculation formula of $\rho(i)$ can be expressed as following formula 9.

$$\rho(i) = \frac{\arctan(i+\alpha)}{\pi} + 0.5 \tag{9}$$

where i is the number of objects with common reviews of two users. Parameter $\alpha \geq 1$ is an adjustment factors of i. The parameter α has a positive integer ranging from [0,5]. D_0 is a default trust value, generally take as $D_0 = \frac{\sqrt{5}-1}{2} \approx 0.618$.

C. PROPAGATION TRUST COMPUTING

According to definition (3), propagation trust of two users can be acquired from direct trust between intermediate users. We proposes a weighted trust graph model G and $G = (U, L, D)$ for all users who review items in E-commerce systems. In the trust graph model, U is the collection of users, denoted as nodes in the graph; L is the set of links that connect nodes; D is the weight of direct trust between the user nodes that are directly connected by L. Let l_{ij} be the path distance between two user nodes (i and j), D_{ij} is the direct trust degree between the two user nodes and set $l_{ij} = -\ln D_{ij}$. According to the definition (3), there are $0 < D_{ij} \leq 1$, $-\ln D_{ij} \geq 0$, and the path distances between all nodes is a set of nonnegative numbers. When the shortest path of two users i and j is calculated, the propagation trust between the users can be obtained by $D_{i,j} = e^{-l_{i,j}}$.

Take part of a weighted trust graph model (as shown in Figure 3) as an example, if $D_{i,k} = 0.6$, $D_{k,j} = 0.8$, then path distances $l_{i,k} = -\ln 0.6$, $l_{k,j} = 0.8$, and

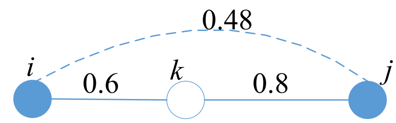


FIGURE 3. Part of a weighted trust graph model.

Algorithm 1 Propagation Trust Calculation

Begin
Initialize:
let $q = \{u_0\}$; $L_0 = 0$; $Q_0 = q$;
//q is a set of source nodes, and it is u_0 at beginning. L is path length. Q is a temporary variable for distance marks.
$L_i = \infty$; $K = \{u_0\}$;
// Select the first node, the distance between the same node is 0, the parent node of the node is itself, the distances between different nodes are set to infinite; K is the node set of the shortest path passes through.
let $Q_i = 0$ for all $i \in H$, $i \neq 0$; //H is the number of user nodes.
let $Y = \{q\}$; $F = \{q\}$;
while $K \neq \varphi$ then do;
// Search for a collection of nodes F , which is within the shortest distance to q;
$u_a = \{i : L_i \leq L_j; i, j \in F, i \neq j\}$
$F = F - \{u_a\}$;
$Y = Y \cup \{u_a\}$;
//F is a temporary variable and Y is the set of nodes that have found the shortest path. Find the shortest path between the nodes u_a and u_b, put the node u_a into set Y;
for each (u_a, u_b), If $L_{u_a} + l_{u_a,u_b} < L_{u_b}$ do;
$L_{u_b} = L_{u_a} + l_{u_a,u_b}$;
//l_{u_a,u_b} is path length of direct link from the set Y that include u_a *to* u_b
$Q_{u_b} = u_a$; //Update the new distance mark;
if $u_b \notin F$ then $F = F \cup \{u_b\}$;
//Update the parent node;
endif;
if $u_a \in K$ then $K = K - \{u_a\}$;
endif;
endwhile;
end

$l_{i,j} = -(\ln 0.6 + \ln 0.8)$. The propagation trust between the users i and j can be calculated as

$$PR_{i,j} = e^{-l_{i,j}} = e^{\ln 0.6 + \ln 0.8} = 0.48$$

Following the weight graph theory, any two user nodes can be acquired by adding up direct trust weight between all nodes in the shortest trust path. We put forward an improved shortest trust path algorithm to calculate the maximum propagation trust, and the specific idea of the algorithm explained as follows algorithm 1.

In algorithm (1), q is a set of source nodes, u_i is a set of user nodes, Q_i is a set of parent nodes for the *ith* node, Y is

TABLE 1. Reviews data structure.

Field	Description
reviewerID	ID of the reviewer
asin	ID of the product
reviewerName	name of the reviewer
helpful	helpfulness
reviewText	text of the review
overall	rating of the review
summary	summary of the review
unixReviewTime	time of the review
reviewTime	time of the review

TABLE 2. Product metadata structure.

Field	Description
asin	ID of the product
title	name of the product
price	price in US dollars
imUrl	url of the product image also-viewed/
related products	bought sales rank information (related to also bought, also viewed, bought together, buy after viewing)
salesRank	brand name
brand categories	list of categories the product

the shortest path set that is already known, and l is the distance between two different nodes.

With the algorithm for the shortest paths, propagation trust of every two users i and j can be acquired. However, if the value of propagation trust l_{ij} is smaller than a threshold parameter ε, that is the degree of trust between them is so small that it does not support trust relationships, We think there is no trust relationship between the two users. The ε is a prior parameter obtained through experiment.

VI. EXPERIMENT EVALUATION

A. DATASET COLLECTION

The experimental dataset is collected from Amazon.com. It contains 143.7 million reviews on 24 categories of products with 9.4 million items from May 1996 to July 2014 [48], [49]. The dataset includes basic information about reviewers, such as user name, user ID, and object information of the review, e.g. the product name or the product ID, the specific text of the review and the user's rating status, etc. Specifically, the dataset contains two files:

(1) review data file, which consists of ratings, text and helpfulness votes (detailed structure of the file is shown in Table. 1).

(2) product metadata file that contains item description, product category information, price, brand, and image features and also-viewed / also-bought (detailed structure of the file is shown in Table. 2).

The whole dataset is more than 350GB, and we just collect reviews on five categories (books, electronic, sports and outdoors, video games and baby) of products as test data for our experiments. The basic statistic information of the selected data is shown in Table. 3.

TABLE 3. Selected dataset of reviews from amazon.

Category	Reviews	Items
Books	22.5M	2.37M
Electronics	7.82M	498K
Sports and Outdoors	1.32M	532K
Video Games	3.26M	51K
Baby	915K	71.3K

B. MEASURES

According to the data description in Tab. 1 and 2, we use three fields: related, overall and helpfulness, which is in the original experiment dataset, to judge whether the trust exists between users. If two users have the same related products (also bought, also viewed, bought together, buy after viewing), the same overall value (rating of the product) and the same helpful value (helpfulness rating of the review), we assume there is trust relationship between them. To our best knowledge, there are no published papers adopt the same or similar approach to explore users' trust from E-commerce system reviews based on sentiment similarity analysis. Therefore, we just follow the existing research and introduce four evaluation indicators, *Precision, Recall, F value* and *Accuracy* to evaluate the effectiveness of our approach [50].

If the trust of any two nodes obtained by the sentiment mining algorithm is greater than a pre-set non-zero threshold value and trust relationship exists actually at the same time, then we deem the result of calculation is correct, and otherwise it is incorrect. If the results acquired by the proposed algorithm do not show trust links (including direct and propagation) relationship between two users, but the trust between them is true actually, this situation is called trust relationship missing. *Precision* and *recall* are defined as following formula 10:

$$precision = \frac{CorrectLinksNumber}{CorrectLinksNumber + IncorrectLinksNumber} \tag{10}$$

$$recall = \frac{CorrectLinksNumber}{CorrectLinksNumber + MissedLinksNumber} \tag{11}$$

According to the relationship between *precision* and *recall*, *F value* is used to represent the common impact of the two indicators on the results and it is defined as formula 12:

$$F - value = \frac{2pecision \times recall}{pecision + recall} \tag{12}$$

In the original experiment dataset, the purchase records of all users and their assessment of the product are generally objective existence, such as if {<also bought> and <also viewed> or <bought together> or <buy after viewing>} and {<same rating value>} and {<same helpfulness rating value of the review} is true, then the two users can be considered to trust each other. therefore, this trust can be used to evaluate the accuracy of the trust which is based on sentiment similarity calculations.

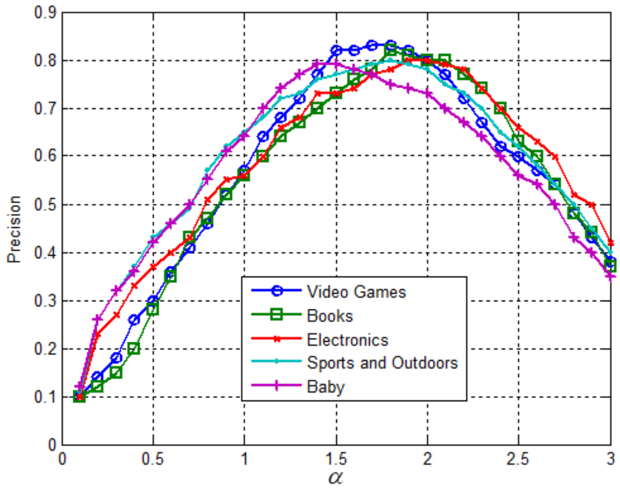


FIGURE 4. The *precision* of direct trust analysis with different α.

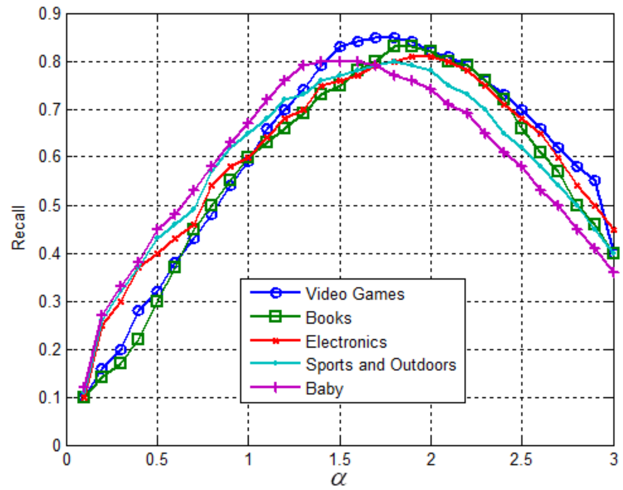


FIGURE 5. The *recall* of direct trust analysis with different α.

The evaluation index can be used to evaluate the trust links relationship found by the sentiment analysis algorithm; that is, the validity of the algorithm is expressed by the ratio of the correct trusted links. The accuracy of the algorithm is defined as formula 13:

$$accuracy = \frac{CorrectLinksNumber}{TotalActualLinksNumber} \tag{13}$$

C. EXPERIMENTAL RESULTS AND ANALYSIS

We divide the review dataset into two parts, which is randomly divided into two non-overlapping parts, namely, training dataset and test dataset. We use 80% of the dataset for training and extracting entity-sentiment word pairs, computing sentiment similarity and trust. We use the left 20% of the dataset for experiment of accuracy and effectiveness of direct and propagation trust.

Both direct trust relationship and the propagation trust are analysed experimentally to get trust between users. In the direct trust experiment, we use the control factor to test different accuracy of direct trust with different value of α. The range of α in our experiment is set to [0, 3].

The experiment of *precision, recall* and *F value* on different α are shown in figure 4, figure 5 and figure 6. These figures indicate that the factor of α plays a significant role in all results. Specifically, the values of *precision, recall, and F Value* grow apparently with the increase of α in the range of [0.1, 1.75]. Meanwhile, with the value of α increase in the range of [1.75, 3], the values of the three indictors decrease obviously.

The weights of ρ along with the increase of α, trust is different with distinct categories, but the overall trends are alike no matter α in range of [0.1, 1.75] or [1.75, 3]. These imply that the weights of ρ are important for the trust accuracy. Because the value of ρ increases monotonically with the value of α, when the value of α increase within a certain range, the accuracy of trust also increases, but when a certain maximum value is reached, the accuracy decreases

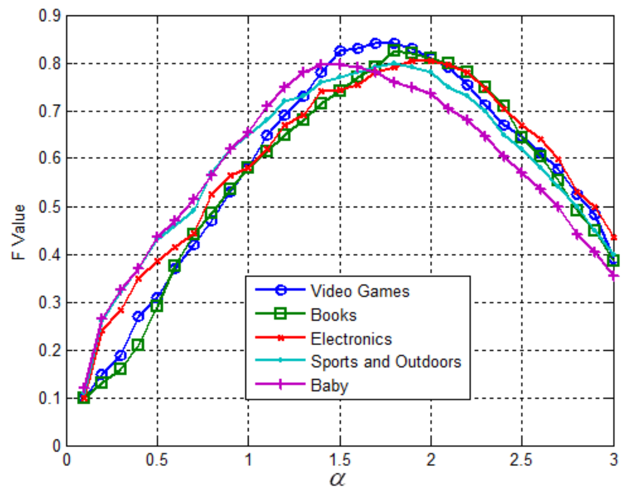


FIGURE 6. The *F-Value* of direct trust analysis with different α.

conversely. Therefore, it is wise to consider the proportionality of the factor α comprehensively.

At the same time, it can be seen from the figures 4 to figure 6 that the influences of α on distinct categories of commodity vary apparently. This reveals the distinguish between the tendencies towards trust and people's views of different items. In addition, when people view reviews on various categories of products, their recognitions are different: opinions on certain commodities, such as Video Games are relatively consistent, and Books, while on others, e.g. Baby Goods, vary widely.

Training sample dataset plays a very important role in entity-sentiment word pair mining that we utilize in the experiment. We conduct an experiment to find the relationship between the number of training samples and the accuracy of trust between users. According to figure 4 to figure 6, when the value of α is close to 1.5, the accuracy is higher. Therefore, we set $\alpha = 1.5$ to perform the accuracy under different

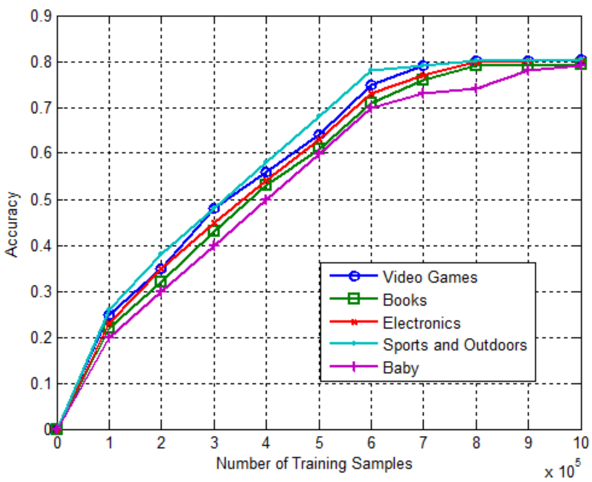


FIGURE 7. Comparison of the effects of different sample numbers on direct trust accuracy.

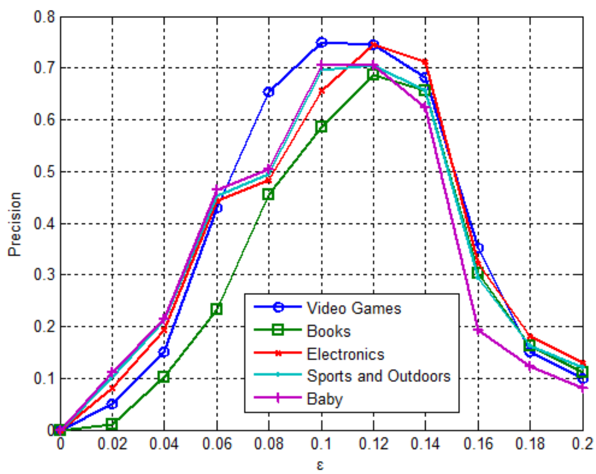


FIGURE 8. The *precision* of propagation trust analysis with different ε.

number of reviews samples. The result is shown in figure 7. We can find from it that the accuracies of five different kind of commodities (video games, books, electronics, sport and outdoors, and baby) increase nearly linearly while the training sample number is in $[0.7 \times 10^5]$. But the accuracies of the five categories of commodities raise slightly or keep the same with the training sample number in $[7 \times 10^5, 10 \times 10^5]$. Total number of the selected training dataset contains more than 10×10^5 reviews, and this ensures the reliability of our experiments.

In propagation trust experiment, the trust path length between every two nodes is controlled by a threshold ε. ε is a value obtained through experiences. The larger the ε, the shorter trust path length is allowed, and the less nodes can be included in the trust path. Whereas smaller ε value means relatively lower sentiment similarity values can be deemed as the existence of trust between users and long trust paths are accepted in algorithm 1.

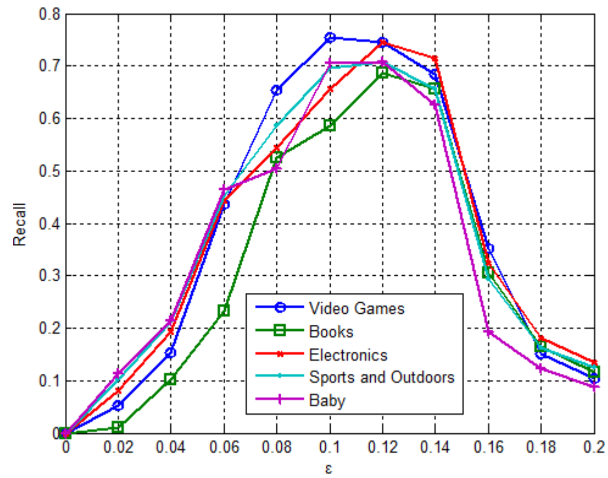


FIGURE 9. The *recall* of propagation trust analysis with different ε.

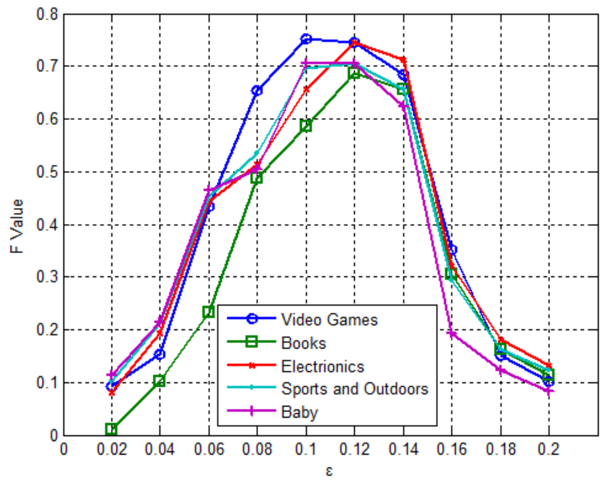


FIGURE 10. The *F-Value* of propagation trust analysis with different ε.

Figure 8 to figure 10 imply a close relationship between the *precision, recall, F value*, and the ε values. If the threshold ε is too small or too large, the experimental results of *precision, recall*, and *F value* are relatively low, indicating that the accuracy of propagation trust calculation algorithm is low. With the increase of ε value in [0.02, 0.1], *precision*, *recall* and *F value* raise obviously. Meanwhile, the values of the three indicators drop sharply with the increase of ε value in [0.12, 0.2]. When ε is set to a relatively low value, for example at 0.02, it means that the propagation trust between nodes can pass through less highly trusted nodes. As a result, the precision on video games, books, electronics, baby, sports and outdoors are very low (as shown in figure 8). But if ε is set to a too large value, e.g. 0.2, propagation trust propagation trust between nodes can pass through a highly trusted only, this will lead to low precision too. At the same time, for diverse types of goods, the best threshold range is also different. Take video games as an example, the best range of ε is [0.1, 0.12], and for electronics, the best range is [0.12, 0.14].

Choosing the appropriate threshold value ε is an important part in improving the propagation trust links analysis. But in real E-commerce system, the lengths of the trust path are different when people make trust judgment on diverse types of commodities. For example, some people prefer to consider a great many pages of reviews before making the final purchase decision, while others just like to reference to a few reviews. Therefore, it is really a challenge to find a perfect ε value.

VII. CONCLUSION AND FUTURE WORK

In our work, we address the problem of mining users trust in E-commerce system. By defining two kinds of trust relationship, namely, direct trust and propagation trust, we transfer the point of exploring trust between users into calculation of sentiment similarity of their reviews. With the help of entity-sentiment word pairs mining, sentiment similarity of reviews can be calculated and direct trust relationships can be obtained through sentiment similarity analysis, which contains of sentiments and ratings aspect. These two aspects can be used jointly to analyse the sentiment direct trust relationship. We establish a weighed trust graph model for propagation trust computing. Propagation trust is the use of the propagation characteristics of trust. It is an indirect trust between two users without direct trust and is obtained through intermediate users who have direct trust between these two users. The propagation trust calculation approach is based on the improved shortest path algorithm, and the time complexity is $O(V^2)$, where V is the number of node in the graph. Ways to improve the computational complexity of the algorithm is a new problem that needs further study because the relatively large number of users in modern e-commerce system.

It can be said with certainty that the user's trust relationship can be obtained through the similarity of them. But the user's trust is not simple normal linear with the user's similarity. How to accurately describe this relationship will be the focus in further research. At the same time, there are several valuable study of sentiment similarity and trust in E-commerce field in the future: (1) Not each user gives their reviews on each item, so the user's reviews data are usually sparse for a particular item. how to explore similarity of users with extremely sparse reviews data, e.g. by designing more efficient algorithm to overcome the challenge; (2) The degree to which people trust others is different for different things. Under more stringent requirements, it is also necessary to distinguish the categories of trust targets in details. how to include other information, for example, purchase item category, brand and other activities, into user sentiment calculation framework and (3) how to incorporate temporal factors to capture users' similarity change will be the focus of future research.

REFERENCES

[1] H. Liu, F. Xia, Z. Chen, N. Y. Asabere, J. Ma, and R. Huang, "TruCom: Exploiting domain-specific trust networks for multicategory item recommendation," *IEEE Syst. J.*, vol. 11, no. 1, pp. 295–304, Mar. 2017.

[2] P.-Y. Hsu, H.-T. Lei, S.-H. Huang, T. H. Liao, Y.-C. Lo, and C.-C. Lo, "Effects of sentiment on recommendations in social network," in *Electron Markets*. Berlin, Germany: Springer, 2018, pp. 1–10, doi: 10.1007/s12525-018-0314-5.

[3] C. Qin, W. Siyi, and A. Lin, "The joint beta distribution with refund rate in online C2C trust building: A theoretical study on Taobao," in *Proc. Int. Conf. E-Learn. E-Technol. Educ. (ICEEE)*, Lodz, Poland, Sep. 2012, pp. 191–196.

[4] S. Kraounakis, I. N. Demetropoulos, A. Michalas, M. S. Obaidat, P. G. Sarigiannidis, and M. D. Louta, "A robust reputation-based computational model for trust establishment in pervasive systems," *IEEE Syst. J.*, vol. 9, no. 3, pp. 878–891, Sep. 2015.

[5] P. De Meo, E. Ferrara, D. Rosaci, and G. M. L. Sarné, "Trust and compactness in social network groups," *IEEE Trans. Cybern.*, vol. 45, no. 2, pp. 205–216, Feb. 2015.

[6] M. G. Ozsoy and F. Polat, "Trust based recommendation systems," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Niagara Falls, ON, Canada, Aug. 2013, pp. 1267–1274.

[7] L. Sheugh and S. H. Alizadeh, "A fuzzy aproach for determination trust threshold in recommender systems based on social network," in *Proc. 9th Int. Conf. E-Commerce Developing Countries, Focus E-Bus. (ECDC)*, Isfahan, Iran, Apr. 2015, pp. 1–5.

[8] Y. Ruan, L. Alfantoukh, and A. Durresi, "Exploring stock market using twitter trust network," in *Proc. IEEE 29th Int. Conf. Adv. Inf. Netw. Appl.*, Gwangiu, South Korea, Mar. 2015, pp. 428–433.

[9] C.-N. Ziegler, *Social Web Artifacts for Boosting Recommenders*, vol. 487. Berlin, Germany: Springer, 2013.

[10] C.-N. Ziegler and J. Golbeck, "Investigating interactions of trust and interest similarity," *Decis. Support Syst.*, vol. 43, no. 2, pp. 460–475, 2007.

[11] N. Lathia, S. Hailes, and L. Capra, "Trust-based collaborative filtering," in *Trust Management II*, vol. 263, Y. Karabulut, J. Mitchell, P. Herrmann, and C. D. Jensen, Eds. Boston, MA, USA: Springer, 2008, pp. 119–134.

[12] D. H. Alahmadi and X.-J. Zeng, "ISTS: Implicit social trust and sentiment based approach to recommender systems," *Expert Syst. Appl.*, vol. 42, no. 22, pp. 8840–8849, 2015.

[13] M. Koufaris and W. Hampton-Sosa, "The development of initial trust in an online company by new customers," *Inf. Manage.*, vol. 41, no. 3, pp. 377–397, 2004.

[14] H. Zhong, S. Zhang, Y. Wang, and Y. Shu, "Study on directed trust graph based recommendation for E-commerce system," *Int. J. Comput. Commun. Control*, vol. 9, no. 4, pp. 510–523, 2014.

[15] M. Li, Y. Xiang, B. Zhang, and Z. Huang, "A sentiment delivering estimate scheme based on trust chain in mobile social network," *Mobile Inf. Syst.*, vol. 2015, Sep. 2015, Art. no. 745095, doi: 10.1155/2015/745095.

[16] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision Support Syst.*, vol. 43, no. 2, pp. 618–644, 2007.

[17] B. Zhang, R. Yong, M. Li, J. Pan, and J. Huang, "A hybrid trust evaluation framework for E-commerce in online social network: A factor enrichment perspective," *IEEE Access*, vol. 5, pp. 7080–7096, 2017, doi: 10.1109/ACCESS.2017.2692249.

[18] D. Alahmadi and X.-J. Zeng, "Improving recommendation using trust and sentiment inference from OSNs," *Int. J. Knowl. Eng.*, vol. 1, no. 1, pp. 9–17, 2015.

[19] Y.-J. Li and Y.-F. Dai, "Research on trust mechanism for peer-to-peer network," *Chin. J. Comput.*, vol. 33, no. 3, pp. 390–405, 2010.

[20] B. Zhang, Y. Xiang, and Q. Xu, "Semantics based information trust computation and propagation algorithm for semantic Web," in *Proc. 5th Int. Conf. Wireless Commun., Netw. Mobile Comput.*, Beijing, China, Sep. 2009, pp. 1–4.

[21] Y. Sun, Z. Han, and K. J. R. Liu, "Defense of trust management vulnerabilities in distributed networks," *IEEE Commun. Mag.*, vol. 46, no. 2, pp. 112–119, Feb. 2008.

[22] P. H. Andersen and R. Kumar, "Emotions, trust and relationship development in business relationships: A conceptual model for buyer–seller dyads," *Ind. Marketing Manage.*, vol. 35, no. 4, pp. 522–535, 2006.

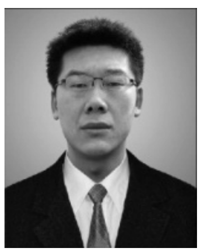
[23] T. DeWitt, D. T. Nguyen, and R. Marshall, "Exploring customer loyalty following service recovery: The mediating effects of trust and emotions," *J. Service Res.*, vol. 10, no. 3, pp. 269–281, 2008.

[24] F. Alam and A. Paul, "A computational model for trust and reputation relationship in social network," in *Proc. Int. Conf. Recent Trends Inf. Technol. (ICRTIT)*, Chennai, India, Apr. 2016, pp. 1–6.

[25] S. Li, I. Doh, and K. Chae, "Non-redundant indirect trust search algorithm based on a cross-domain trust model in content delivery network," in *Proc. 19th Int. Conf. Adv. Commun. Technol. (ICACT)*, Bongpyeong, South Korea, Feb. 2017, pp. 72–77.
[26] F. Yuan, H. Li, and Q. Fu, "A new trust propagation and aggregation method for P2P networks," in *Proc. 2nd Int. Conf. Instrum., Meas., Comput., Commun. Control*, Harbin, China, Dec. 2012, pp. 1176–1181.
[27] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
[28] S. Y. Ganeshbhai and B. K. Shah, "Feature based opinion mining: A survey," in *Proc. IEEE Int. Advance Comput. Conf. (IACC)*, Banglore, India, Jun. 2015, pp. 919–923.
[29] R. Ren, D. D. Wu, and T. Liu, "Forecasting stock market movement direction using sentiment analysis and support vector machine," *IEEE Syst. J.*, to be published, doi: 10.1109/JSYST.2018.2794462.
[30] J. Kikuchi and V. Klyuev, "Gathering user reviews for an opinion dictionary," in *Proc. 18th Int. Conf. Adv. Commun. Technol. (ICACT)*, Pyeongchang, South Korea, Jan./Feb. 2016, pp. 566–569.
[31] T. Takehara, S. Miki, N. Nitta, and N. Babaguchi, "Extracting context information from microblog based on analysis of online reviews," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, Melbourne, VIC, Australia, Jul. 2012, pp. 248–253.
[32] K. Mouthami, K. N. Devi, and V. M. Bhaskaran, "Sentiment analysis and classification based on textual reviews," in *Proc. Int. Conf. Inf. Commun. Embedded Syst. (ICICES)*, Chennai, India, Feb. 2013, pp. 271–276.
[33] J. Akaichi, "Social networks' Facebook' statutes updates mining for sentiment classification," in *Proc. Int. Conf. Social Comput.*, Alexandria, VA, USA, Sep. 2013, pp. 886–891.
[34] F. Colace, M. De Santo, and L. Greco, "Sentiment mining through mixed graph of terms," in *Proc. 17th Int. Conf. Netw.-Based Inf. Syst.*, Salerno, Italy, Sep. 2014, pp. 324–330.
[35] H.-Y. Kao and Z.-Y. Lin, "A categorized sentiment analysis of Chinese reviews by mining dependency in product features and opinions from blogs," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol.*, Toronto, ON, Canada, Aug./Sep. 2010, pp. 456–459.
[36] L. Yang, Y. Zhang, C. Xing, and T. Zhang, 'A node interest similarity based P2P trust model," in *Proc. IEEE 12th Int. Conf. Commun. Technol.*, Nanjing, China, Nov. 2010, pp. 572–575.
[37] H. Zhang, Y. Wang, and X. Zhang, "Transaction similarity-based contextual trust evaluation in E-commerce and E-service environments," in *Proc. IEEE Int. Conf. Web Services*, Washington, DC, USA, Jul. 2011, pp. 500–507.
[38] G. Wang, X. Gui, and G. Wei, "A recommendation trust model based on E-commerce transactions content-similarity," in *Proc. Int. Conf. Mach. Vis. Hum.-Mach. Interface*, Kaifeng, China, Apr. 2010, pp. 105–108.
[39] J. Golbeck, "Trust and nuanced profile similarity in online social networks," *ACM Trans. Web*, vol. 3, no. 4, pp. 1–30, 2009.
[40] H. Mohammadhassanzadeh and H. R. Shahriari, "Using user similarity to infer trust values in social networks regardless of direct ratings," in *Proc. 9th Int. Conf. Inf. Secur. Cryptol. (ISC)*, Tabriz, Iran, Sep. 2012, pp. 66–72.
[41] M. R. Garakani and M. Jalali, "A trust prediction approach by using collaborative filtering and computing similarity in social networks," in *Proc. Int. Congr. Technol., Commun. Knowl. (ICTCK)*, Mashhad, Iran, Nov. 2014, pp. 1–4.
[42] Q. Zhang, Y. Zhuo, and Z. Gong, "A trust inspection model based on society behavior similarity rule in dynamic networks," in *Proc. Int. Conf. Comput. Sci. Softw. Eng.*, Hubei, China, Dec. 2008, pp. 970–973.
[43] G. Cai, M. Wang, Y. Wang, and Y. Zhou, "A collusion detection trust model based on behaviour similarity," in *Proc. IET Int. Conf. Inf. Commun. Technol. (IETICT)*, Beijing, China, Apr. 2013, pp. 241–245.
[44] K. Khan and B. B. Baharudin, "Analysis of syntactic patterns for identification of features from unstructured reviews," in *Proc. 4th Int. Conf. Intell. Adv. Syst. (ICIAS)*, Kuala Lumpur, Malaysia, Jun. 2012, pp. 165–169.
[45] Z. Zhai, H. Xu, J. Li, and P. Jia, "Sentiment classification for Chinese reviews based on key substring features," in *Proc. Int. Conf. Natural Lang. Process. Knowl. Eng.*, Dalian, China, Sep. 2009, pp. 1–8.
[46] K. Bloom and S. Argamon, "Automated learning of appraisal extraction patterns," *Lang. Comput.*, vol. 71, no. 1, pp. 249–260, 2009.
[47] S. Robertson, "Understanding inverse document frequency: On theoretical arguments for IDF," *J. Document.*, vol. 60, no. 5, pp. 503–520, 2004.
[48] J. McAuley, C. Targett, Q. Shi, and A. Van Den Hengel, "Image-based recommendations on styles and substitutes," in *Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, Santiago, Chile, Aug. 2015, pp. 43–52.
[49] J. McAuley, R. Pandey, and J. Leskovec, "Inferring networks of substitutable and complementary products," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Sydney, NSW, Australia, Aug. 2015, pp. 785–794.
[50] L. Havrlant and V. Kreinovich, "A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation)," *Int. J. General Syst.*, vol. 46, no. 1, pp. 27–36, 2017.

SHAOZHONG ZHANG was born in Jinzhou, Liaoning, China, in 1969. He received the B.S. degree in computer science from the University of Shenyang Jianzhu, Liaoning, China, in 1991, the M.S. degrees in world economy from the University of Nankai, Tianjing, China, in 1999, and the Ph.D. degree in computer science from the Dalian University of Technology, Dalian, China, in 2004.

From 2006 to 2008, he was a Postdoctoral Researcher with Beihang University, Beijing, China, and also a Senior Visiting Scholar with Dalarna University, from 2015 to 2016. Since 2008, he has been a Professor with the College of Electronic and Computer Science, Zhejiang Wanli University, China. He has authored two books and over 80 articles. His research interests include data mining and knowledge discovery, sentiment analysis, big data mining, and E-commerce and business intelligence.



HAIDONG ZHONG was born in Wuhan, Hubei, China, in 1982. He received the B.S. degree in computer science and technology, in 2004, and the Ph.D. degree in cartography and geography information system from East China Normal University, in 2011. He is currently an Associate Professor with the College of Logistics and E-commerce, Zhejiang Wanli University. At present, his main research interests include mobile e-commerce, personalized recommendation, and business intelligence.