E-Commerce Reviews Users Trust Analysis Based on Sentiment Similarity and LIWC

Wen Li, Zhouhan Yu, Baoheng Huang, Chi Feng, Mark Newman

2019-08-08

1 Abstract

Online customer review system has developed for 20 years to help consumers to record and share their opinions on their purchased items or services. Many studies and surveys have shown more than 90% of shoppers check reviews at least once before hitting the online buy button. Customer reviews system is playing an important role of providing recommendations for prospective customers to make decisions. In social science, a vast of existing studies researched on trust relationship phenomenon between strangers and many researchers have proved users are willing to trust reviews who have similar aspects. The success of our referred study from Zhang and Zhong has demonstrated the use of sentiment similarity analysis can be an efficient way to discover trust between users in E-commerce systems (Zhang & Zhong, 2019). The purpose of this study is to measure and compare the relation between sentiment similarity and user trust. We propose a new method to calculate the sentiment similarity with LIWC (Linguistic Inquiry and Word Count), which can extract different psychological and linguistic features from natural language text. We show that the LIWC method puts more weight on the emotion of the review texts, while the current studies focus on analyzing sentiment aspect of reviews more statistically. Although LIWC is powerful to extract the tone of the customer review text, our research results reveal users trust relationships can only be partially obtained through sentiment similarity

Keywords: Sentiment similarity, LIWC, trust, E-commerce reviews

2 Introduction

Customer Reviews, which are posted by customers, are they related to the feedback of purchased items or services. Generally, the reviews can reflect users' experience. The e-commerce system is akin to buying and selling products and services online or over the internet. Typically, customers generally pay attention to the reviews when they want to purchase products in online shopping sites. Since 1999, the online reviews began showing up for the process of execution of commercial transactions. With the sustainable evolution of technology, media, networking, the increasing number of people are willing to leave their opinion and suggestions based on the items or services to someone who has some interest in them. People generally have the ideas which can be shared by friends or relatives, or check the reviews posted online by a stranger if they want to be served on the Web. Multicategory item recommendation, which is based on the reviews, can be applied to decision making(Hsu et al., 2019). Reviews with a negative polarity classification received more attention than those with a positive polarity classification(Chen, Wang, & Lin, 2012).

Meanwhile, the review mining-based applications are useful to make people's decisions. With the development of society, the behavior patterns of human beings change substantially, and the applications paired with the patterns consistently(Liu et al., 2017). For example, there is an increasing amount of people who are willing to buy products, book a flight or car online, they usually check the item related to customer reviews first, not consult with friends. Consistently, numerous national/international E-commerce companies construct customer reviews systems, such as Amazon, BestBuy, Walmart, and Department stores.

When someone has a desire for decision making, the online experience can do much favor. In other words, people need to trust others with a precise method. Generally, the trust in friendship and family can help to get the information, detail, opinion, suggestion for item or service chosen that are not familiar. However, there is an apparent phenomenon that people tend to trust strangers over computer networks(Kraounakis et al., 2015). In the electronic market, a risk level associated with the loss of the notions of trust and reputation is raised. Although the e-commerce has been developed rapidly as a commercial trade pattern, unlikely customers have no chance to conversation face to face with the seller or agent or enjoy the direct experience in-store with the associate.

Meanwhile, the customer services, include return, exchange, refund and repair services, will be delayed for days or weeks compared the services in stores. Thus, online customer reviews are essential to product demanders. Online review posters can share the experiences and have some suggestions about related online services with buyers.

Nevertheless, each review system covers massive scale feedback, which can reflect detailed historical records from all reviewers related to each product. The overall review texts, which can be trusted, is challenging to differentiate by the consumers. Generally, individuals hold abnormal standpoint and field to view one item. In this case, they hold personalized interests, opinions which differ by each other over the items or services. For example, some people gave a positive review of the car because of the color while others did the negative one due to function. In this case, the customer group has a different main point in the vehicle. At this point, no review text can be available for each person and attitude.

In social science, a vast of papers researched on this specified trust relationship phenomenon between strangers. A finding was shown up: users are willing to trust reviews who have similar aspects as much as possible. The essential factors can include review text pattern, sentiment words, items, and age period(Ozsoy & Polat, 2013). It is challengeable to figure out the relationship between users and similarity customers.

As the researchers, research can use sentiment analysis to find and measure customer attitudes to items, services — for example, LIWC (Linguistic Inquiry Word Count). LIWC can provide features from we calculated linguistic similarity. Typically, this paper would apply it to determine the sentiment similarity group and compare the results with general calculation.

The objective of this paper is to create a new approach for sentimental similarity, which can be applied to calculate the direct or propagation trust in an e-commerce scenario. In the following part of this paper is categorized as follows. In the literature review section, this paper describes the history and related works for this research, which included trust computation studies, sentiment, and similarity analysis issues, LIWC sentiment analysis, correlations between trust and similarity, problems and our solutions. In the method section, there are a specified approach and detail steps on calculations of sentiment similarity, direct trust, and propagation trust. Meanwhile, researchers notified related algorithm definitions, the framework of the trust relationship based on sentiment similarity. In the analysis section, the paper demonstrated the results based on the algorithm as described and made a comparison between them and LIWC results. Also, it analyzed the comparison and concluded that. In the conclusion section, the paper summarized the findings and discussed with possible points, which are to be furthered.

3 Literature Review

3.1 Trust Computation

Many researchers had been worked on defining the concept of trust in different areas. In economy filed, the trust has a rational explanation, and the action happened after the careful consideration and cost calculation. In other words, individuals tend to make sensible and profit-maximizing choices, which is often referred to as computational trust in existing research(Liu et al., 2017).

Direct trust and indirect trust (M. Liu & Huang, 2015) are the core of trust computation. Direct trust can be built by observing the interaction relationship between the observations. Dimah explores the potential of social information as a source of user-related recommendations. They suggest a way to take advantage of two factors in online social networks: the emotional orientation of a friend's posts about a particular project and the trust relationship between friends. Li and Dai (LI & DAI, 2010) proposed a promising method for dealing with P2P network trust mechanism. They let the parties rate each other after the transaction is completed and use the aggregate ratings of a given party to generate trust scores.

However, indirect trust is widely used to connect users with long paths through intermediate users (Jøsang, Ismail, & Boyd, 2006). This trust is transferable from a familiar context to a new context, or from a trusted entity to an unknown one. Bo, Yang, and Qiang proposed an information transmission mechanism based on the semantic network for the calculation of semantic trust score. The whole trust is measured by the comprehensive trust score from both subjective and objective information (Zhang, Xiang, & Xu, 2009). The actual aspect of trust is the semantic trust of information, and the personal element is the trust relationship between peers. The trust relationship is based on the historical interaction between peers (Sun, Han, & Liu, 2008) and (DeWitt, Nguyen, & Marshall, 2008). Faruk and Arnab put forward a trust management model, which takes direct trust, indirect trust, global trust, and other factors into account to obtain the ultimate trust value of services. The trust path distance will be considered when the research calcuates indirect trust. Li et al. proposed two trust models of a content delivery network: local trust model and cross-domain trust model. Hong (Yuan, Li, & Fu, 2012) proposed a new method, Max-aggregation, which calculates the reputation of peers through multidimensional and multi-attribute perspective and obtains indirect reputation through trust propagation and aggregation in the trust network.

3.2 Sentiment Similarity Analysis

The sentiment similarity analysis method takes similarity analysis as its primary content and takes emotion and emotion as the evaluation factors of trust. Besides, in natural language understanding, data mining, and statistical analysis (Zhang & Liu, 2017), emotion and emotion similarity analysis have also been extensively studied.

The existing emotional analysis methods based on similarity mining can be divided into three levels: document level, sentence level, and entity feature level. These three levels are all based on the opinion dictionary. Opinion dictionary is a set of specific keywords or emotional words (extracted from collected comments), with part-of-speech labels, which serve as the basis for analyzing the reliability of comments (Kikuchi & Klyuev, 2016). At the document level, the task is to classify the whole opinion document according to whether it expresses positive emotions or negative emotions (Takehara, Miki, Nitta, & Babaguchi, 2012) and (Mouthami, Devi, & Bhaskaran, 2013). At the sentence level, the task considers the sentence and determines whether each sentence shows a positive, negative, or neutral view. Neutral usually means that no opinions are given (Akaichi, 2013), (Colace, Santo, & Greco, 2014). Neither at the document level nor the sentence level can analysis find out exactly which particular objects people like or dislike. At the entity and feature levels, this approach focuses directly on the idea itself, rather than looking at the language structure (document, paragraph, sentence, clause, or phrase). It is based on a view that consists of emotions (positive or negative)(Kao & Lin, 2010). Xu (Hsu et al., 2019) used emotion word database to extract emotion-related data from posts and used these data to study the influence of different types of emotion words on product recommendation.

The primary purpose of existing emotion analysis methods is to cluster users' emotions, which are usually divided into several types. Even at the entity level and feature level, the main purpose is to classify users' emotions into likes and dislikes. However, the above method is directly related to the overall trend, which is not enough when we calculate trust based on emotional similarity. In the comments, it is necessary to analyze specific attitudes towards specific objects.

3.3 The Correlation between Trust and Similarity

Over the past decades, there was considerable research focused on the relationship between trust and similarity. Emotional similarity analysis has become an important research method to build a trust relationship. Many studies have shown that there is a high correlation between trust and similarity. They show that similar people also have high levels of trust in interests, contents, behaviors, etc.

Ziegler and Golbeck have studied the relationship between trust and similarity of interests. They established a formal framework to study the interaction between trust and similarity. They used a mathematical model to calculate the similarity and proposed the calculation algorithm of profile and profile similarity. At the same time, the analysis of Film Trust website data shows that when the user similarity changes within a specific range, the trust between users also changes. This change indicates that there is a strong relationship between trust and similarity. Li proposed a trust model based on node interest similarity, which takes into account the interest preference of nodes and reputation of each interest domain, and USES reputation vector of interest domain to maintain the behavior of nodes in specific interest domain. The local trust recommendations (Li Yang, Yusen Zhang, Changyou Xing, & Tao Zhang, 2010) is weighted by the interest similarity between each node. These creative studies demonstrate the correlation between trust and similarity and propose corresponding calculation methods.

3.4 Linguistic Inquiry and Word Count

LIWC is a text analysis software that can enable the extraction of different psychological and linguistic features from natural language text. LIWC is a kind of natural Language Processing (NLP) technology, which can conduct quantitative analysis of text content and calculate different categories of words (primarily psychological words) of text files leading to people, such as the percentage of psychological words such as causal words, emotional words and cognitive words in the whole text .ll.After more than ten years of development, modification and expansion, LIWC has become increasingly stable. After LIWC,LIWC2001 and LIWC2007, the LIWC mainly consists of two parts: the program body and the dictionary. Among them, the core is the dictionary, which defines the category name and word list of words and phrases. The program compares the words in the text with the dictionary by importing the dictionary and text, and outputs the word frequency results of different words. The current LIWC contains four general descriptive categories (total number of words, number of words per sentence, more than six-letter words, crawl rate), 22 language feature category (such as personal pronouns, and auxiliary verbs, conjunctions, prepositions), 32 psychological feature categories (such as social process, emotional process words, cognitive processes, physiological processes, etc.), 7 personalization categories (such as work, leisure, family, money, etc.), three pair of linguistic categories (e.g., with words, pause wordage, fill the eulogy, etc.) and 12 punctuation categories (such as a full stop, comma, colon, semicolons, etc.), a total of 80 - word category, about 4500 words(Tausczik & Pennebaker, 2010).

As a tool to measure the psychological characteristics of language, LIWC is bound to be tested by reliability and validity. In general, when a self-reported questionnaire subject design, can is a specific psychological characteristic compiled several different situation or subject description way, and, assuming that these questions are able to reflect the same psychological characteristics, then the consistency between subject evaluation should be quite high, and on the basis of calculating coefficient of internal consistency. A, usually in 0.7 above can be an excellent internal consistency. However, the use of language is different from the self-assessment questionnaire. We do not deliberately alternate synonyms every time we write or talk (for example, we do not deliberately alternate synonyms of “I” and “I” or “myself”). In two writings or conversations on the same topic, we may · changes the emphasis and is less likely to use the same words and content (for example, two self-introductions may be too rigid if the content is the same). Therefore, language analysis tools cannot be used to measure internal consistency and reliability according to the standard of general questionnaires. Nevertheless, the English version of LIWC2007 tested for internal consistency. The internal consistency of all categories ranges from 0.02 to 0.75. Except for a few relatively low categories, the internal consistency of most categories is within the acceptable range.

Mainly, the validity test of LIWC is realized through criterion validity. For example, people with depression use more negative words and more first-person singular nine pronouns. It is also found that the more positive emotional words used, the more negative emotional words used, and the more causal words and Epiphany words used, the healthier the writers are.

4 Method and Materials

4.1 Data Description

Amazon reviews data are downloaded from http://jmcauley.ucsd.edu/data/amazon/. It includes 11.3 million product reviews from five categories.The review texts are collected from Amazon.com, and reviews are dated from May 1996 to July 2014. The datasets we used in the analysis are the 5-core datasets, which are the subset of the whole dataset. 5-core means that all users and items have at least five reviews. The detials can be checked in Table 1.

Categories	**Reviews**
Books	5-core (8,898,041 reviews)
Electronics	5-core (296,337 reviews)
Sports and Outdoors	5-core (296,337 reviews)
Video Games	5-core (231,780 reviews)
Baby	5-core (160,792 reviews)

Table 1: Amazon 5-core Review data

Amazon reviews data contain nine fields, and in this research, there are six fileds which are listed in Table 2.

Columns	**Description**
reviewerID	ID of the reviewer
asin	ID of the product
helpful	helpfulness
reviewText	text of the review
overall	rating of the product
reviewTime	time of the review

Table 2: Review Data Contents

Note that, due to the size of the datasets, Baby category data is selected as the testing dataset to replicate the whole research. Most of the researches are coded in Python. The measurement steps may take too much time to process in Python. Therefore, that step is completed in Bigquery.

4.2 LIWC with Amazon Review Data

All the reviews data are also run through LIWC. The results contain different sentiment scores from different perspectives for each review. The sentiment similarity in this research and the sentiment scores in LIWC seem to be different from each other. Our team uses LIWC results in three approaches to make sure we make the comparison in an effective way.

First, use the LIWC tone column as the final trust scores. Second, calculate the Pearson Correlation among all the common reviews' tone values between two users and use the tone correlation scores as the final trusts scores. Third, use all the fields from LIWC and calculate ordinary reviewed object's Pearson Correlation. Use that as the sentiment similarity scores on the same object between two users. The rest calculation steps are the same as the pipeline in the study.

4.3 Trust Calcuation Procedure

The computing framework involves four steps in general (Zhang & Zhong, 2019). First, to extract entity-sentiment word pairs from each review text and calculate the mutual information (MI) score of an entity-sentiment word pair. Second, to calculate two review texts' sentiment similarity based on the entity-sentiment word pairs. The two review texts should come from two different users and be left for the same object. Third, to aggregate the universal objects' similarity scores between two users and calculate a direct trust score between two users. Last, to estimate the propagation trust, the indirect trust, for users have no common reviewed objects based on the direct trust scores. Details of each step will be explained in the following paragraphs.

4.3.1 Step 1: Entity-Sentiment Word Pair and Mutual Information

The researchers define all the nouns or noun phrases as entity words; therefore, entity words may include the objects, and some related features and attributes. The detection of the nouns and noun phrases is completed through python package NLTK's part-of-speech (POS) tagging, which is the same function for tagging the sentiment words. There are two sub-steps to detect sentiment words. On the one hand, the research considers all the adjectives and adverbs as the entity words. On the other hand, the study includes all the sentiment words from WordNet lexicon. In our methods, we did POS tagging with wordnet initiated, and all the sentiment words are tagged with "Positive", "Neutral", and "Negative" for the sake of the later adjusted function.

Not all the entity and sentiment words will become a valid entity-sentiment word pair. Only the word pairs, first, existing at the same time in the same paragraph of the review text, at the same time, having their MI score over a certain threshold, will be saved as the basis for the next step's calculation.

The MI algorithm in the research is TF-IDF (Term Frequency-Inverse Document Frequency)(Robertson, 2004), which is expressed in formula 1.

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

In the equation above, $p(e_i)$ represents the occurrence frequency of entity words e_i, $p(s_i)$ represents the occurrence frequency of sentiment words s_i, and $p(e_i, s_i)$ is the joint occurrence frequency of the entity-sentiment (E-S) word pairs (e_i, s_i).

The emotional expression of the sentiment words are kept in the previous steps because the research adds an adjusted weight to each entity-sentiment word pair. The adjusted weights $(w^{'})$ can be expressed in formula 2.

$$w^{'}_{i,j} = TF_{i,j} \cdot IDF \cdot \theta(s_{i,j}) = \frac{n_{i,j}}{\sum\limits_{k=1}^{N}} \cdot log\frac{N}{n_k} \cdot [1 + log(1 + \frac{n_j}{N} \cdot \lambda(s_j))] \tag{2}$$

In equation 2, N is the paragraph number in a review text, and n_k is the number of paragraphs the word appears. The adjust function here is $\theta(s_j)$. The primary factor in the adjust function is $\lambda(s_j)$. If the sentiment word is not neutral, then $\lambda(s_j) = 1$, otherwise, $\lambda(s_j)$ will be 0.

So far, we have a dictionary of E-S word pairs r and their corresponding adjusted weight scores $w^{'}$, which can be expressed in formula 3.

$$\overrightarrow{r_j} = ((r_1, w^{'}_1), r_2, w^{'}_2), ...(r_i, w^{'}_i)) \tag{3}$$

4.3.2 Step 2: Sentiment Similarity Between Review Texts

The idea behind sentiment similarity in this step is the Pearson Correlation. To calculate the sentiment similarity between two users' reviews of the common objects is to calculate the Pearson Correlation between two sets of MI scores. The two sets have to be the same length; therefore, a subset of word-pairs will be used in the correlation calculation. The similarity formula can be expressed in formula 4.

$$Sim(t_{u_i}, t_{u_j}) = \frac{\sum_{k=0}^{min|V|}(r_{ik} - \overrightarrow{\overline{r_i}}) \cdot (r_{jk} - \overrightarrow{\overline{r_j}})}{\sqrt{\sum_{k=0}^{min|V|}(r_{ik} - \overrightarrow{\overline{r_i}})^2 \cdot \sum_{k=0}^{min|V|}(r_{jk} - \overrightarrow{\overline{r_j}})^2}} \tag{4}$$

In the equation 4, V is the number of E-S pairs in reviews texts between user u_i and u_j.

4.3.3 Step 3: Direct Trust Calculation Between Two Users

The research aggregates all the review sentiment similarity scores between two users of the common objects and finalizes all of them into one direct trust score. The formula can be expressed in formula 5.

$$D_{u_i,u_j} = \frac{1}{O_{u_i} \bigcap O_{u_j}} \cdot (\sum_{i \in O_{u_i} \bigcap O_{u_j}} \rho(i) \frac{\sum_{k=0}^{|T|} \sigma(S_{u_i}, S_{u_j}) \cdot Sim(t_{u_i}, t_{u_j})}{\sum_{i \in O_{u_i} \bigcap O_{u_j}} Sim(t_{u_i}, t_{u_j})} + (1 - \rho(i)) \cdot D_0) \tag{5}$$

In equation 5, $O_{u_i} \bigcap O_{u_j}$ represents the number of objects that two users reviewed at the same time. $\sigma(s_{u_i}, s_{u_j})$ is the influence function here, which is a score related to user ratings and can be calculated through formula 6. S in the following equation represents the rating scores to the familiar objects.

$$\sigma(s_{u_i}, s_{u_j}) = \frac{\sum_{i=1}^{n}(s_{u_i} - \overline{s_{u_i,u_j}}) \cdot (s_{u_j} - \overline{s_{u_i,u_j}})}{n-1}, n \neq 1 \tag{6}$$

The last crucial part of formula 5 is $\rho(i))$. $\rho(i))$ is an adjustment function here for user interaction level α. The parameter i represents the number of common reviewed objects of two users. D_0 is a fixed value here. The range for α is from 0 to 5.

4.3.4 Step 4: Propagation Trust Calculation Between Two Users

The calculation for propagation trust is not hard as long as the shortest path between two "indirect" users is found. The Python package "networkx"can be used to extract all the shortest path here, and the theory behind it is the weighted graph model. The propagation trust can be calculated as formula 8.

$$l_{ij} = -lnD_{ij} \tag{7}$$

$$PR_{ij} = -e^{-ij} \tag{8}$$

Basically, the trust score thresholds are applied to both direct trust and propagation trust, which are parameters that calculated based on the distribution of the direct trust scores and propagation trust scores.

4.4 Measures

Although continuous trust scores are calculated, the assessment of the whole pipeline is based on the confusion matrix. Trust scores between 0.07 and 0.08 are deemed as trust relationship existed between two users. Otherwise, no trust relationship exists. The true trust relationship in the confusion matrix is defined by the researchers. As long as two users reviewed the same objects or related products, at the same time, they had the same overall ratings and helpfulness scores, then the trust relationship between two users is considered as true trust.

For the three LIWC approaches, the theory of trust relationship is similar to the normal process. Take the first approach, for instance. If both two users have a positive or negative attitude towards the same objects, and the tone score differences are within a specific range, then those two users have a trust relationship. It is noted that due to the limitation of LIWC results, our team only uses the first approach as a result comparison in the following analysis.

Three major evaluation indicators are Precision, Recall, and F-Value (Havrlant & Kreinovich, 2017). The formulas for those three indicators are as follows.

$$precision = \frac{CorrectLinksNumber}{CorrectLinksNumber + IncorrectLinksNumber} \tag{9}$$

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

$$F - value = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{11}$$

Since we have the user interaction level as the control function, we have different Precision, Recall, and F-Value scores under different α values. Similarly, since we have different range limitation for the trust relationship, we also have different evaluation values for LIWC results.

5 Results

The goal of the research is to analyze the trust relationship between users statistically. There is a control function α in direct trust, which could affect the results. α is set from 0 to 3, with 0.1 difference at each step. The direct trust's results of Precision, Recall and F-value with different control function α are shown in figure 1, figure 2 and figure 3. As shown in the following graphs, with the increase of the user interaction level α, the control function, the Precision, Recall and F-Value also increase. The peak reaches with α equal to 1.7. After the max values, Precision, Recall and F-values all decrease with the increase of α.

The above graphs suggest that the user interaction level α plays a vital role in trust relationship between users. The Recall figure 2 reflects the distributions of direct trust values under different α values are skewed towards a "central" value, therefore, the drop on both sides is steep.

The LIWC's results from three perspectives are very different from the regular approach in the study. Take the first approach as the example here, and figure 4, figure 5 and figure 6 display the Precision, Recall and F-value of direct trust with different LIWC thresholds. We set the user differences in tone values, that is the thresholds, from 1 to 50. Figure 4 shows that the Precision decreases with the tone difference range increases, while the Recall and F-Value trends are the opposite. The highest point of Precision happens at the beginning of the plots, when the difference threshold is 1. On the opposite, the highest point of Recall and F-Value happen at the very end of the plots when the threshold is 50.

The significant differences between the study's approach and LIWC approaches indicate that either our group have not found the best way to utilize LIWC results, or the researchs approach and LIWC focus on different aspects of sentiment analysis. LIWC put more weights on the emotion of the review texts, while the research is from a more statistical perspective to analyze sentiment aspects of the review texts.

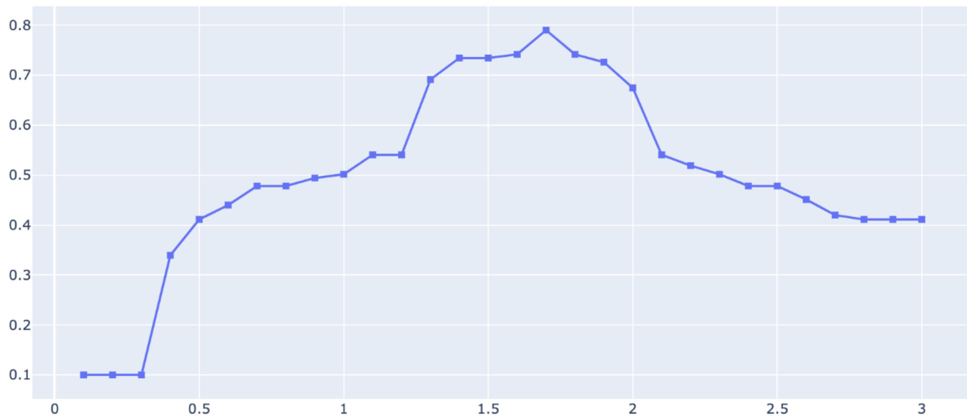

Figure 1: The *precision* of direct trust analysis with different α

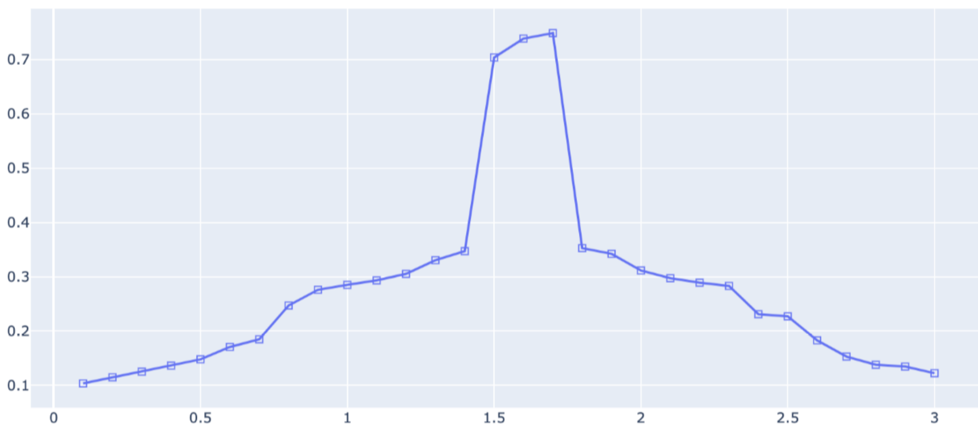

Figure 2: The *recall* of direct trust analysis with different α

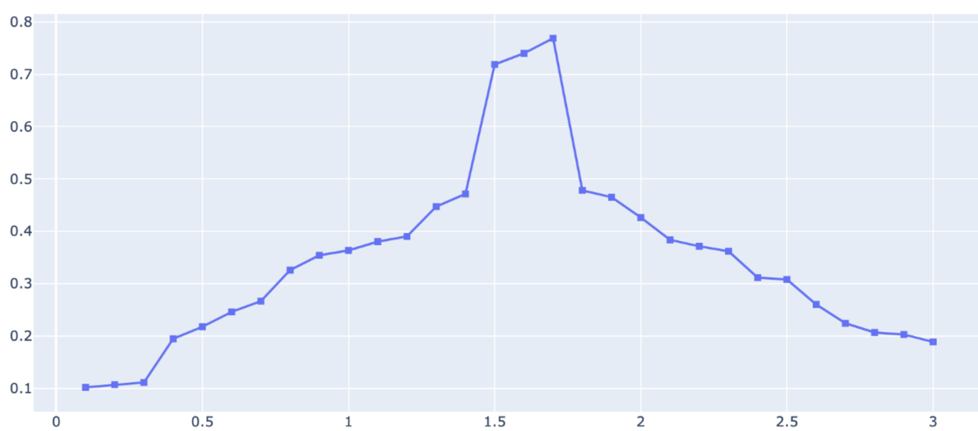

Figure 3: The *F-value* of direct trust analysis with different α

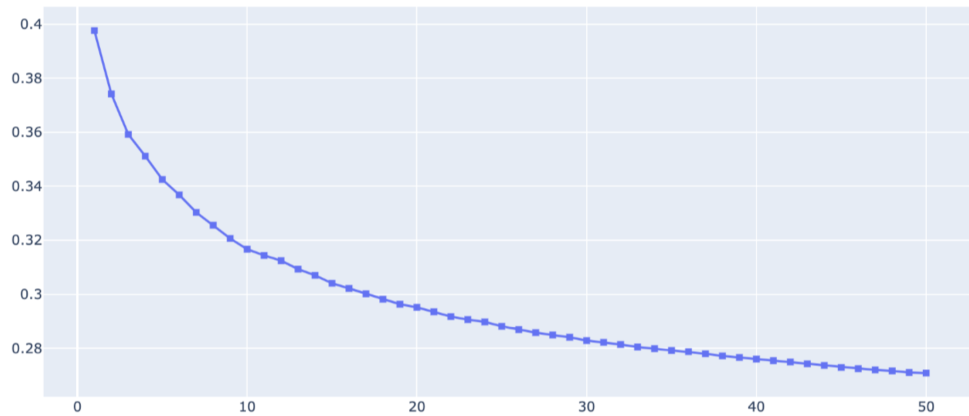

Figure 4: The *precision* of direct trust analysis with different LIWC threshold

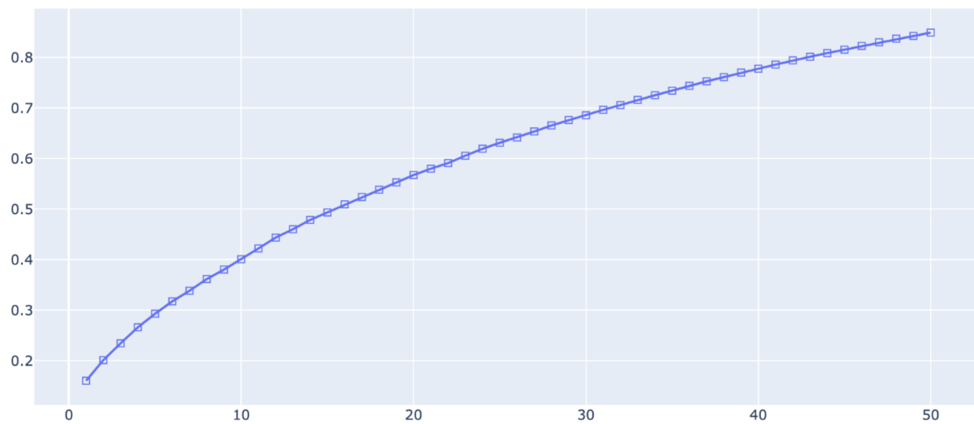

Figure 5: The *recall* of direct trust analysis with different LIWC threshold

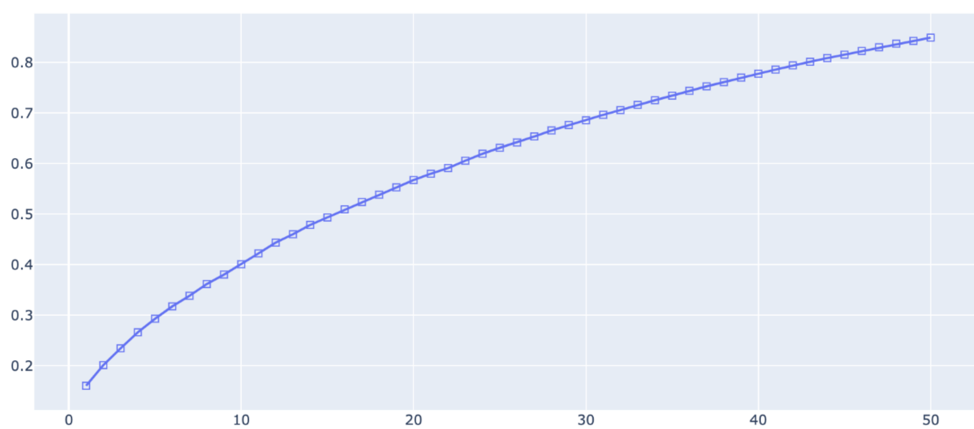

Figure 6: The *F-value* of direct trust analysis with different LIWC threshold

6 Conclusion and Discussion

The purpose of the research is to find the relationship between sentiment similarity and user trust. Based on the methods in the study, we found that the trust relationship between users can be obtained through sentiment similarity among users' reviews. However, LIWC's results tell another story. Our different results from LIWC are telling us that users trust relationships can only be partially obtained through sentiment similarity. There should be other factors playing an important role in trust relationships.

Another question that we have is regarding the true values in the confusion matrix. The researchers define their own trust relationship on common columns with common ratings and helpful scores. If two users meet those conditions, are they going to "trust" each other with regard to online shopping? This definition could affect the results a lot.

All in all, we agree that reviews' sentiment similarity between users plays an essential role in users' trust relationship based on the research's definition. However, if there is a trust definition that can work with both the research's algorithm and LIWC's approach, then we can say for sure that there is an association between reviews sentiment similarity and users trust relationships.

7 References

Akaichi, J. (2013). Social networks' facebook' statutes updates mining for sentiment classification. In *2013 international conference on social computing* (pp. 886–891).

Chen, Q., Wang, S., & Lin, A. (2012). The joint beta distribution with refund rate in online c2c trust building: A theoretical study on taobao. In *2012 international conference on e-learning and e-technologies in education (iceee)* (pp. 191–196).

Colace, F., Santo, M. D., & Greco, L. (2014). Sentiment mining through mixed graph of terms. In *2014 17th international conference on network-based information systems* (pp. 324–330).

DeWitt, T., Nguyen, D. T., & Marshall, R. (2008). Exploring customer loyalty following service recovery: The mediating effects of trust and emotions. *Journal of Service Research*, *10*(3), 269–281.

Havrlant, L., & Kreinovich, V. (2017). A simple probabilistic explanation of term frequency-inverse document frequency (tf-idf) heuristic (and variations motivated by this explanation). *International Journal of General Systems*, *46*, 27–36.

Hsu, P.-Y., Lei, H.-T., Huang, S.-H., Liao, T. H., Lo, Y.-C., & Lo, C.-C. (2019). Effects of sentiment on recommendations in social network. *Electronic Markets*, *29*(2), 253–262. Retrieved from https://doi.org/10.1007/s12525-018-0314-5

Jøsang, A., Ismail, R., & Boyd, C. (2006). A survey of trust and reputation systems for online service provision. *Decision Support Systems*, *43*, 618–644.

Kao, H., & Lin, Z. (2010). A categorized sentiment analysis of chinese reviews by mining dependency in product features and opinions from blogs. In *2010 ieee/wic/acm international conference on web intelligence and intelligent agent technology* (Vol. 1, pp. 456–459).

Kikuchi, J., & Klyuev, V. (2016). Gathering user reviews for an opinion dictionary. In *2016 18th international conference on advanced communication technology (icact)* (pp. 566–569).

Kraounakis, S., Demetropoulos, I. N., Michalas, A., Obaidat, M. S., Sarigiannidis, P. G., & Louta, M. D. (2015). A robust reputation-based computational model for trust establishment in pervasive systems. *IEEE Systems Journal*, *9*(3), 878–891.

LI, Y.-J., & DAI, Y.-F. (2010). Research on trust mechanism for peer-to-peer network: Research on trust mechanism for peer-to-peer network. *Chinese Journal of Computers*, *33*, 390–405.

Liu, H., Xia, F., Chen, Z., Asabere, N. Y., Ma, J., & Huang, R. (2017). TruCom: Exploiting domain-specific trust networks for multicategory item recommendation. *IEEE Systems Journal*, *11*(1), 295–304. Retrieved from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7114205

Li Yang, Yusen Zhang, Changyou Xing, & Tao Zhang. (2010). A node interest similarity based p2p trust model. In *2010 ieee 12th international conference on communication technology* (pp. 572–575).

M. Liu, B. Z., Y. Xiang, & Huang, Z. (2015). A sentiment delivering estimate scheme based on trust chain in mobile social network. *Mobile Information Systems*.

Mouthami, K., Devi, K. N., & Bhaskaran, V. M. (2013). Sentiment analysis and classification based on textual reviews. In *2013 international conference on information communication and embedded systems (icices)* (pp. 271–276).

Ozsoy, M. G., & Polat, F. (2013). Trust based recommendation systems. In *2013 ieee/acm international conference on advances in social networks analysis and mining (asonam 2013)* (pp. 1267–1274).

Robertson, S. E. (2004). Understanding inverse document frequency: On theoretical arguments for idf. *Journal of Documentation*, *60*, 503–520.

Sun, Y., Han, Z., & Liu, K. J. R. (2008). Defense of trust management vulnerabilities in distributed networks. *IEEE Communications Magazine*, *46*(2), 112–119.

Takehara, T., Miki, S., Nitta, N., & Babaguchi, N. (2012). Extracting context information from microblog based on analysis of online reviews. In *2012 ieee international conference on multimedia and expo workshops* (pp. 248–253).

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24–54.

Yuan, F., Li, H., & Fu, Q. (2012). A new trust propagation and aggregation method for p2p networks. In *2012 second international conference on instrumentation, measurement, computer, communication and control* (pp. 1176–1181).

Zhang, B., Xiang, Y., & Xu, Q. (2009). Semantics based information trust computation and propagation algorithm for semantic web. In *2009 5th international conference on wireless communications, networking and mobile computing* (pp. 1–4).

Zhang, L., & Liu, B. (2017). Sentiment analysis and opinion mining. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning and data mining* (pp. 1152–1161). Boston, MA: Springer US. Retrieved from https://doi.org/10.1007/978-1-4899-7687-1_907

Zhang, S., & Zhong, H. (2019). Mining users trust from e-commerce reviews based on sentiment similarity analysis. *IEEE Access*, *7*, 13523–13535.