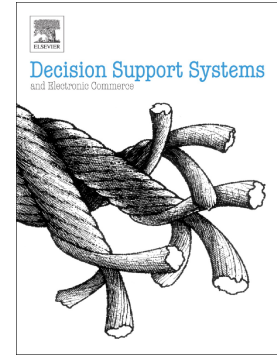# Journal Pre-proof

A text analytics framework for understanding the relationships among host self-description, trust perception and purchase behavior on Airbnb

Le Zhang, Qiang Yan, Leihan Zhang

Please cite this article as: L. Zhang, Q. Yan and L. Zhang, A text analytics framework for understanding the relationships among host self-description, trust perception and purchase behavior on Airbnb, *Decision Support Systems* (2020), https://doi.org/10.1016/j.dss.2020.113288

A text analytics framework for understanding the relationships among host self-description, trust perception and purchase behavior on Airbnb

Le Zhang[a], Qiang Yan[a,*], Leihan Zhang[b]

[a]*School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing 100876, PR China, lezhang920@gmail.com, yan@bupt.edu.cn*
[b]*Wangxuan Institute of Computer Technology, Peking University, Beijing 100871, PR China, zhangleihan@gmail.com*

*Corresponding author
Email address: yan@bupt.edu.cn (Qiang Yan)
URL: `http://sem.bupt.edu.cn/info/1304/6027.htm` (Qiang Yan)

Abstract

Trust plays an important role in sharing transactions on short-term rental platforms. However, the impact of host self-description on trust perception and whether trust perception can influence purchase behavior remain under-studied. Therefore, a text analytics framework was proposed to research the relationships among host self-description, trust perception and purchase behavior on Airbnb. Specifically, a deep-learning-based method was designed to automatically code trust perception of host self-descriptions. And the linguistic and semantic features of description texts were extracted with text mining methods. The estimated order quantity was used to quantify purchase behavior. Then, the influence of linguistic and semantic features on trust perception was identified, and the relationship between trust perception and purchase behavior was also verified. The empirical analysis derives the following findings: i. The readability of self-description is positively associated with trust perception; ii. Perspective taking expressed in self-description is also helpful; iii. Excessive positive sentiment expression can raise barriers to trust building; iv. Paying more attention to family relationship, openness, service and travel experience in self-description would be helpful; v. Trust perception can promote purchases. These findings can help hosts write better self-description, which contributes to trust building and purchases on short-term rental platforms.

Keywords: Self-description, Trust perception, Purchase behavior, Airbnb, Text mining

1 Introduction

The rapid development of information technology and network services has brought about the sharing economy boom, and various kinds of online platforms, such as Airbnb, Uber, and Zipcar, have emerged, enabling people to engage in providing services and sharing idle resources with others [1]. Airbnb, a peer-to-peer (P2P) accommodation sharing platform, is one of the most successful models in the sharing economy, providing short-term rental services for more than 200 million guests across 191 countries [2]. On Airbnb, guests make bookings via the online platform and pay for the accommodation service. A complete transaction usually contains the following steps: (1) the guests search for target units on the website and then attempt to contact the hosts via an instant messaging system; (2) the two parties attempt to build trust through their reputation, self-disclosure and communication; (3) once an agreement is reached, guests can check-in and will likely have some face-to-face interactions with the hosts for a period of time; and (4) finally, the guests check out and post a review about the accommodation experience. For the parties engaged in the transaction, hosts can earn money and make friends by sharing idle homes or rooms with guests, and guests can directly interact with hosts and obtain low-cost accommodations. From this perspective, sharing accommodations on Airbnb can be deemed a mixed-mode interaction of social exchange (i.e., online first, then offline) [3]. However, this kind of social exchange may involve higher risks and greater uncertainty. For instance, guests usually begin their social interactions with expectations of gaining a unique accommodation experience, building social relationships, and understanding different cultures [4]. However, these expectations cannot be guaranteed because the online service is intangible, and the quality of the stay cannot be verified until it is experienced [5]. Moreover, guests are also at risk of facing unreliable hosts or even threats to their personal safety.

Confronted with high risk and uncertainty, guests use whatever information they can obtain, such as online reputation and host profile details, to make trust inferences and purchasing decisions. Among these variables, rating scores and reviews, as components of reputation systems, have been widely researched in traditional e-commerce. Consequently, trust-building is usually simplified to the use of a reputation system [6]. However, trust on Airbnb is far more complicated and does not rely solely on reputation. In fact, the comments on Airbnb are often highly positive [7]. Thus, the reputation system alone may not be enough to identify reliable hosts. Faced with the complex scenario of building trust in the sharing economy, hosts are encouraged to make self-disclosures to reduce guests' sense of uncertainty and risk, such as providing a personal photo or writing a self-description. Thus, whether and how these host attributes can impact guests' trust perception and decision-making has attracted researchers' attention.

The existing literature mainly focuses on the impact of a host's numeric attributes, reputation and profile photo on guests' perceived trust and purchase intention. Some studies have found that host gender, emotion, and attractiveness revealed in personal photos influence guests' perceived trust and purchase intention [8, 9, 10, 11]. Moreover, the superhost badge, response behavior, and verifications can also impact guests' perceived trust [11]. Self-descriptions can reveal more details about hosts than a profile photo. Specifically, hosts can introduce themselves on their personal page and express any topics and sentiments through various writing styles. The numbers of words and topics contained in self-descriptions were found to have a positive influence on guests' perceived trust [12]. Similarly, positive sentiment being expressed in a host's self-description can also help build guests' perceived trust [11]. However, the types of linguistic style and semantic content of host self-descriptions that can help build guests' perceived trust on

Airbnb remain understudied. Moreover, prior studies have often focused on the relationship between guests' trust perception and purchase intention, but little is known about how guests' perceived trust influences purchase behavior.

In this paper, we propose a text analytics framework for understanding the relationships among self-description, trust perception and purchase behavior based on real-world data from Airbnb. A deep-learning procedure was employed to automatically code trust perception. The linguistic features, including readability, sentiment intensity, and perspective taking, were extracted and quantified using text mining methods. The semantic topics hidden in self-descriptions were identified by LDA (latent Dirichlet allocation) [13]. Then, robust regression was applied to identify the impact of linguistic features and semantic features on trust perception. Finally, the positive influence of trust perception on purchase behavior was verified. Here, purchase behavior was quantified by the estimated order quantity. Our study obtained several interesting findings: (1) both readability and perspective taking have a significant positive impact on trust perception, while sentiment intensity has an inverted U-shaped relationship with trust perception; (2) we found that hosts who prefer to express their family relationships, openness, service and travel experience can obtain higher trust perceptions than hosts who mainly express demographic-related information such as age, occupation and personal interests; and (3) we also found that guests' trust perception can promote purchases on Airbnb. These findings can help hosts improve their self-descriptions, which contributes to trust building and purchase behavior on short-term rental platforms.

The remainder of this paper is organized as follows. Section 2 summarizes related studies. In Section 3, the data and methods are introduced. Section 4 presents the results and analysis. In section 5, the findings, conclusions, implications, and limitations of the work are discussed.

2. Literature review

2.1. Trust in the sharing economy

The sharing economy is an economic phenomenon that enables the public to obtain income without changing ownership by sharing under-utilized resources [14]. In contrast to traditional e-commerce, the sharing economy pays more attention to providing services such as short-term rental services and ride-hailing services, and the parties involved in sharing transactions usually need to meet with one another face-to-face. Consequently, the quality of these services is usually difficult to guarantee [5]. Moreover, the self-interest of peers is not limited to economic rewards but also involves their psychological needs [15]. Hence, the sharing economy can be conceptualized in terms of reciprocity, risk, and interdependence. The self-interest of peers engaged in sharing transactions cannot be realized without the cooperation of their partners, and risk is an inescapable part of this process.

In traditional B2C or C2C e-commerce, providers mainly focus on selling products, but consumers are not able to see or touch the products prior to purchase, and customer service is also limited. Thus, consumers usually perceive trust and make purchase decisions according to provider reputation, which are mainly assessed by rating scores and comments [8]. Naturally, researchers mainly concentrate on the influence of provider reputation on building consumer trust and reaching an agreement, where reputation plays a pivotal role in building trust [16]. Inspired by this consensus, nearly all online sharing platforms integrate reputation mechanisms into their systems. However, it is difficult for guests to judge a host's trustworthiness based solely on online reputation. This is because the reviews on Airbnb often contain a limited set of linguistic

resources, and almost all of the reviews are highly positive, which makes the reviews seem very similar on the surface [17]. For instance, nearly 95% of Airbnb listings have an average guest-generated rating higher than 4.5 stars (the maximum is 5 stars), and nearly none of the listings have a rating less than 3.5 stars [7]. In fact, in addition to reputation, many other factors can also influence guests' trust.

On Airbnb, the guests browse shared accommodations with various expectations of economic benefits, diverse living experiences, building social relationships and learning about different cultures, which means that the guests usually endeavour to obtain a comprehensive understanding of the hosts they are dealing with and infer the trustworthiness of the hosts accordingly. Therefore, it is necessary to investigate how various types of information about hosts can influence the trust perception of guests. Many trust-related findings have been disclosed on Airbnb. The "superhost" badge on Airbnb is an important indicator of host reputation, which can help improve perceived trust [6]. Hosts who respond to guests more quickly are often considered more hospitable and trustworthy [10]. The verifications of a host can provide his personal identity and indicate that the host does exist, which can also help build trust [4]. The facial emotion and gender revealed in profile photos can also affect guests' trust and purchase intention [8, 9]. Ma [12] found that the number of words and topics in self-descriptions can reduce the uncertainty and positively influence the perceived trust of guests. The above-mentioned factors influencing perceived trust can generally be divided into four kinds: reputation, performance, self-disclosure, and social presence cues [18]. Among the four kinds of factors, self-disclosure probably contains the most abundant information. For example, profile photos can directly reveal a person's face, gender, and even age. The self-descriptions can convey information about hobbies, work, values, and even personality [12]. With this information about hosts, guests can likely build a relatively comprehensive understanding of the hosts, which can further help reduce uncertainty [19] and build trust [11, 12]. From the perspective of hosts, crafting a better self-description can reduce the cost of building trust and obtain more guests. However, to the best of our knowledge, the existing research on self-descriptions basically remains at the level of words and phrases, and deeper features such as linguistic style and semantic content remain under-researched. Considering the diverse needs of guests on Airbnb, we believe that self-description is a valuable proxy for guests to understand hosts and make trust inferences. Thus, we contend that it is worth conducting a deep and comprehensive analysis of the influence of self-description on guests' trust perception.

2.2. Trust, purchase intention and purchases behavior

Trust is the basis of social connection and acts like a bridge to link strangers [20]. Whether in economic exchange or social exchange, trust plays an important role in facilitating successful exchanges. In e-commerce marketplaces, serious information asymmetry problems exist between consumers and providers, and consumers face the risk of information leakage and economic loss [21]. Sharing economy platforms have higher demands for trust than e-commerce marketplaces because consumers not only face information leakage and the risk of economic loss but also have personal safety risks [8]. Thus, trust becomes a prerequisite for consumers to support their purchase decisions.

Purchase intention refers to the willingness of a customer to buy a product or service under certain conditions [22]. By contrast, purchase behavior is a choice made by customers in a real purchasing context, and it is clearly influenced by various factors, such as host attributes, product attributes, reputation, or assurance [23]. In addition, since purchase behavior contains a series of

purchase occasions, the relationship between purchase intention and purchase behavior is nonlinear [24]. It has been found that 90% of consumer behavior is unconscious [25]. Namely, what consumers say could be inconsistent with what they actually do. Thus, purchase intention is different from purchase behavior.

Many studies on P2P short-term rental have focused primarily on the impact of perceived trust on guests' purchase intention [3, 8, 9, 12] rather than purchase behavior in a real decision-making context. Thus, little is known about whether guests' trust perception can influence purchase behavior on short-term rental platforms. In this work, we investigate the relationship between trust perception and purchase behavior on Airbnb.

Figure 1: The process of the research method

3. Dataset and method

The text analytics framework employed to understand the relationships among host self-description, trust perception and purchase behavior is shown in Figure 1. The framework consists of three main parts: data engineering, feature extraction and data analysis. In general, we first collected and cleaned the experimental data. A deep-learning-based method was applied to automatically code the value of trust perception. Then, the linguistic features (e.g., readability, sentiment intensity, perspective taking) and semantic features (e.g., topics) were obtained by using text mining methods. Third, robustness regression analysis was applied to identify the effect of linguistic features and semantic features on trust perception. Finally, the positive influence of trust perception on purchase behavior was verified through negative binomial regression analysis.

3.1. Data engineering

3.1.1. Data collection and preprocessing

To reduce the interference of cultural and temporal bias, we analysed a full year of active listings of 4 cities, which are from North America, Europe and Asia. In detail, we collected 293,948 active listings in New York, London, Paris and Hong Kong from Inside Airbnb[1], and the time range is from October 2018 to September 2019. This raw dataset contained public information about hosts and listings on Airbnb (i.e., self-description, superhost status, verified status, response time, response rate, price, listing type, cancellation policy).

Since our focus was on host self-description texts, we preprocessed the self-description texts through the following steps: (1) removed listings with empty self-description text field; (2) filtered the listings with repeated self-description texts; (3) detected and removed non-English texts by using langdetect[2]; and (4) dropped the listings with self-description text whose word count (not including punctuation marks) was less than five words. Then, we obtained 73,654 listings. The details on the listings for each city are shown in Table 1.

---

[1] http://insideairbnb.com/get-the-data.html
[2] https://pypi.python.org/pypi/langdetect

Table 1: Details of the dataset

City	Continent	Total listings	Selected listings
New York	North America	66,941	23,476
London	Europe	87,591	30,327
Paris	Europe	80,598	10,762
Hong Kong	Asia	58,818	9,089
All		293,948	73,654

Notes: Time period is from October 2018 to September 2019.

3.1.2. *Trust perception coding*

The majority of related research has conducted experiments or designed questionnaires to let participants rate trust perception [3, 8, 12, 26], which is time-consuming and limited in size. We developed a deep-learning procedure to automatically code a trust perception score for each host self-description based on the attention-based bidirectional long short-term memory model (AB-BiLSTM), which can readily be applied to a large amount of data.

Neural network models have been widely used in the field of natural language processing. Recurrent neural networks (RNNs) are capable of processing sequences with arbitrary lengths, and there has been incredible success in applying RNNs to a variety of problems, such as language modelling, translation, and series prediction. As a special kind of RNN, long short-term memory networks (LSTMs) [27] can not only learn long-term dependencies but also avoid the problem of vanishing gradients of RNNs. Additionally, the bidirectional LSTM (BiLSTM) can learn bidirectional long-term dependencies between time steps of sequence data. The attention mechanism [28] enables BiLSTM to focus on a subset of sequence data. Therefore, the attention-based bidirectional LSTM (AB-BiLSTM) was employed to build a trust prediction model and the structure of the model is modified from Zhou [29]. As shown in Figure 2, the AB-BiLSTM model contains five components: (1) Input layer: input sentences to the model, (2) Embedding layer: represent each word with a low dimension compact vector; (3) BiLSTM layer: use BiLSTM to get high level features from the output of embedding layer; (4) Attention layer: produce a weight vector, and merge word-level features from each time step into a sentence-level feature vector, by multiplying the weight vector; (5) Output layer: the sentence-level feature vector is finally used for perceived trust prediction.

Figure 2: The architecture of the AB-BiLSTM model

We used the public human-annotated dataset[3] offered by Ma [26] as the training data. The dataset contains 4,179 English-only unique host self-descriptions from 12 U.S. cities, and the trust perception scores of these self-descriptions were annotated by volunteers recruited on Amazon Mechanical Turk according to a six-item perceived trust scale based on dimensions of integrity, benevolence and ability [30]. We trained the AB-BiLSTM on the 4,179 labeled data and then used the learned model to predict 73,654 unlabelled self-descriptions. In detail, we used the Skip-gram (Word2Vec) model [31] to pre-train the word embedding on both the labeled and unlabelled self-descriptions. The length of word embedding was set to 200, and the number of iterations was set to 8. The architecture of the proposed AB-BiLSTM can be seen in Figure 2. The

---

[3] https://github.com/sTechLab/AirbnbHosts-Extended

numbers of neutron units in the BiLSTM layers were 256. The maximum length of sequences was 200. The batch size was 128, and the dropout probability was 0.5. The self-descriptions were split into a training set, a development set and a testing set at a ratio of 3:1:1. Then, we trained a prediction model using AB-BiLSTM and compared its performance with the method of Ma [26]. Specifically, they used the LIWC (linguistic inquiry and word count) [4] features, sentence categories, and word count to predict perceived trust with the regression method. The LIWC features were extracted using LIWC, which is a dictionary-based text analysis tool. Sentence categories indicate the eight categories manually developed by Ma [12]. Word count refers to the number of words in host self-description. Finally, the explained variance score (EVS), mean absolute error (MAE), mean squared error (MSE) and R^2 were used to evaluate the performance of the AB-BiLSTM-based prediction model. The results shown in Table 2 suggest that the proposed prediction model outperforms the method of Ma [26] on all the indicators. Thus, it is reliable to predict the trust perception score of each self-description using the trained model.

Table 2: Performance of the prediction model

Methods	EVS	MAE	MSE	R^2
AB-BiLSTM	0.597	0.351	0.209	0.558
The method of Ma [26]	0.408	0.401	0.270	0.408

3.2. Feature extraction

3.2.1. Linguistic style

The linguistic style refers to the expression style of verbal content and the text features of word choice, personal pronouns, and adjectives [32], which can imply the writer's personality and psychological needs and can further affect readers' perceived trust [33]. On Airbnb, the linguistic style of hosts' self-descriptions can influence the perceived trust of potential guests towards the hosts from the aspects of reading difficulty, emotional expression and perspective taking. Therefore, we derived the following three linguistic features: readability, sentiment intensity and perspective taking.

(1) Readability. Readability can help improve the understanding of textual information [34]. It can be used to measure the amount of work required for readers to understand a piece of text [35]. Lower readability will create barriers to comprehending the information in the text. Conversely, higher readability makes it easier and faster for readers to absorb information in the text. Furthermore, it can also enhance the persuasive effect of self-descriptions. Here, the Flesch formula [36] was used to measure text readability. For a given text, the greater the Flesch score, the easier it is to understand. Thus, the readability of description texts can be quantified as follows:

$$Flesch\ score = 206.835 - 1.015 * \frac{total\ words}{total\ sentences} - 84.6 * \frac{total\ syllables}{total\ words} \tag{1}$$

(2) Sentiment intensity. Prior research has found that positive sentiment expression by service providers can engender greater customer trust [37, 38]. Similarly, expressing positive

---

[4] https://liwc.wpengine.com/

sentiment in self-description can also help build trust [11, 12]. However, excessive positive sentiment expression can have adverse consequences [39]. It has been found that higher positive sentiment expression can lead to higher perception of warmth but lower perception of service ability [40]. Service ability can also influence consumers' trust perceptions [30]. In addition, when readers encounter texts with high sentiment intensity, they may feel that the writers are attempting to persuade them by exaggerating or even lying [41]. Thus, we believe that sentiment intensity in host description texts may have an inverted U-shaped relationship with trust perception. The positive sentiment being expressed in self-descriptions can help improve guests' trust perception. However, when the sentiment intensity is beyond a point, the trust perception will be reduced.

Here, we proposed a method to calculate the sentiment intensity of self-description texts based on Stanford Sentiment TreeBank (SST)[5] [42]. First, each self-description text was split into sentences according to punctuation such as periods, exclamation points and question marks. Second, SST was used to classify the sentences into five categories according to their sentiment strength: neutral, positive, very positive, negative, and very negative. Third, the sentiment intensity of the above five sentiment categories was set to 0, 1, 2, -1 and -2. Finally, the sentiment intensity of each text was calculated by the average sentiment intensity of all sentences. The formula is shown below:

$$Sens(d_i) = \frac{\sum_{s_j \in d_i} eSens(s_j)}{N} \qquad (2)$$

$Sens(d_i)$ represents the sentiment intensity of self-description d_i. s_j is the sentence in self-description d_i, and $eSens(s_j)$ is the sentiment intensity of s_j. N is the total number of sentences contained in text d_i.

(3) Perspective taking. Perspective taking is the process by which one person views a situation from the perspective of another person [43], and it can suggest the ability of a person to comprehend and take on the viewpoint of another person's psychological experience, such as thoughts, feelings and attitudes [44]. It is acknowledged that perspective taking can lead to beneficial results for interpersonal communication [45]. For instance, a negotiator thinking more about what would satisfy the other party will increase the probability of reaching a deal [46], and reviews posted by consumers with higher perspective taking ability tend to be trusted and adopted by other potential consumers [47]. Thus, we contend that perspective taking is beneficial for earning trust in the sharing economy. Since the use of personal pronouns in a text could provide information about the focus or the subject of attention and has been reliably linked to perspective taking [48], we used the formula 3 proposed by Peng [48] to measure perspective taking (PT) in self-descriptions.

$$PT = \frac{2ndPersonPronouns}{1stPersonPronouns + 2ndPersonPronouns + 0.0001} \qquad (3)$$

3.2.2. *Semantic content*

On Airbnb, the semantic topics presented in self-descriptions can provide more detailed

---

[5] https://nlp.stanford.edu/sentiment/treebank.html

information about hosts, such as their family, occupation, and personality [12]. This information can help guests build a relatively comprehensive understanding of the hosts, which can further help reduce uncertainty [19] and build trust. It has been found that more informative self-descriptions tend to earn higher trust perceptions and the information about interactions and services in self-descriptions is benificial [11, 12]. In fact, the topics in self-descriptions cover a very wide range and the different roles of these topics play in building trust are still not clear. Thus, we believe that it is necessary to find an efficient and reliable method to extract topics from self-descriptions and perform a thorough and comprehensive analysis of the influence of the topics on trust perception.

Table 3: Topics of host self-descriptions identified by using LDA

T1 Personal traits	T2 Family	T3 Openness	T4 Hobby	T5 Work
friendly 0.064	family 0.048	forward 0.079	love 0.061 0.017	I'm 0.049
going 0.05	husband 0.032	looking 0.034	cooking 0.035	work 0.035
easy 0.049	children 0.024	world 0.030	travelling 0.031	designer 0.030
person 0.044	wife 0.020	people 0.029	reading 0.027	fashion 0.025
clean 0.038	love 0.016	guests 0.024	music 0.023	artist 0.024
I'm 0.027	married 0.016	great 0.017	cycling 0.010	creative 0.023
respectful 0.024	couple 0.013	welcoming 0.016	food 0.010	photographer 0.015
tidy 0.014	daughter 0.012	meeting 0.016	movies 0.010	director 0.014
quiet 0.014	mother 0.011	enjoy 0.012	yoga 0.009	music 0.013
professional 0.013	son 0.010	staying 0.010	walking 0.009	freelance 0.011
T6 Local expert	T7 Accommodation	T8 Service	T9 Origin	T10 Travel
city 0.034	guests 0.023	stay 0.061	I'm 0.031	love 0.046
local 0.031	property 0.019	make 0.051	living 0.029	place 0.031
places 0.029	experience 0.016	feel 0.038	years 0.027	enjoy 0.027
restaurants 0.020	apartments 0.014	help 0.034	originally 0.019	world 0.025
visit 0.014	service 0.010	happy 0.033	born 0.016	great 0.013
best 0.014	hotel 0.009	questions 0.024	moved 0.015	beautiful 0.020
offer 0.014	provide 0.008	free 0.023	raised 0.015	experience 0.016
museums 0.009	accommodation 0.007	welcome 0.020	lived 0.013	travel 0.013
bars 0.009	estate 0.007	comfortable 0.020	neighbourhood 0.012	amazing 0.012
enjoy 0.008	hospitality 0.007	ask 0.017	native 0.011	wonderful 0.011

Notes: T1 Personal traits: the personality and character traits of hosts. T2 Family: family members and relationships. T3 Openness: the openness to communicate and make friends with guests from different countries or cultures. T4 Hobby: favourites of hosts. T5 Work: current or past job of hosts. T6 Local expert: the competency of hosts in taking care of guests, offering information about the residence or local culture, e.g., restaurants, museums and bars. T7 Accommodation: information concerning the accommodation. T8 Service: the accommodation services that the hosts can provide for the guests. T9 Origin: residence, life experiences and personal history. T10 Travel: travel experiences and favourite places.

LDA [13] is used to identify topics hidden in self-description texts. Let $D = \{d_i \mid i \in [1, M]\}$ represents a set of host self-descriptions, d_i represents a self-descrition and M is the number of self-descriptions. Given D, LDA can learn the topic set $T = \{t_j \mid j \in [1, N]\}$, the word set $W = \{w_k \mid k \in [1, L]\}$, the distribution set of topics on each self-description $\Theta = \{\theta_i \mid i \in [1, M]\}$ and the distribution set of words on each topic $B = \{\beta^k \mid k \in [1, N]\}$ from the self-description texts. Here, N is the number of topics, L is the number of words, t_j represents a topic, w_k represents a word, $\theta_i = \{(t_j, p_{ij}) \mid j \in [1, N]\}$, p_{ij} represents the probability of t_j in d_i, $\beta_j = \{(w_k, q_{jk}) \mid k \in [1, L]\}$, q_{jk} represents the probability of w_k in t_j. In other words, for each host self-description, we can get the distribution of a number of topics and each topic is represented by a group of words with weight, which represents the size of contribution to this topic. Since the self-descriptions may contain many possible trivial topics, we determine the appropriate number of topics by refering to topic coherence [49]. Then, the topic number of LDA is set to 20. And the selected 10 topics were manually summarized as "Personal traits", "Family", "Openness", "Hobby", "Work", "Local expert", "Accommodation", "Service", "Origin" and "Travel", and the detailed representation of each topic is shown in Table 3. For each topic, the corresponding value of each word indicates the probability that the word belongs to the topic. For each self-description d_i and the distribution $\theta_i = \{(t_j, p_{ij}) \mid j \in [1, N]\}$ of topics in d_i, p_{ij} will be used to represent the weight of topic t_j in d_i.

4. Data analysis and results

4.1. The effect of textual features on trust perception

We performed robust regression to investigate the effect of linguistic features and semantic features on trust perception. The use of robust regression ensures that our model will not be overly affected by violations of underlying assumptions. Compared with widely used methods of regression, such as ordinary least squares, robust regression is less likely to be impaired by outliers, heteroscedasticity and non-normality [50].

The number of notional words of self-description was used to represent information amount, which has been found to be associated with trust perception [11, 12]. Therefore, we employed it as a control variable. We also controlled for a city variable. Then, we checked the collinearity between variables using the variance inflation factor (VIF). As shown in Table 4, there is no evidence of serious multi-collinearity in these data [51]. At last, these variables were normalized by z-scores before the regression analysis.

Table 4 reports the regression results. Model-1 incorporated only the control variables. Model-2 analysed the influence of linguistic features on trust perception in conjunction with the control variables. Model-3 added the squared term of sentiment intensity. The semantic features of ten topics were added in Model-4. As Model-4 is the mixed model and contains all the variables, the results were mainly analysed based on Model-4.

Table 4: Results of robustness regression analysis

2*Variables	Dependent variable: trust perception			
	Model-1	Model-2	Model-3	Model-4
Control variables				
#Notional words	0.075***	0.061***	0.060***	0.047***
	(0.002)	(0.002)	(0.002)	(0.001)
City_HongKong	0.047***	0.027***	0.027***	-0.006
	(0.005)	(0.005)	(0.005)	(0.005)
City_NewYork	-0.067***	-0.063***	-0.065***	0.002
	(0.004)	(0.004)	(0.004)	(0.004)
City_Paris	-0.123***	-0.145***	-0.138***	-0.088***
	(0.005)	(0.005)	(0.005)	(0.004)
Independent variables				
Readability		0.022***	0.016***	0.007***
		(0.002)	(0.002)	(0.001)
Sentiment intensity		0.019***	0.038***	0.032***
		(0.002)	(0.002)	(0.002)
Sentiment intensity ^{2}			-0.052***	-0.052***
			(0.002)	(0.002)
Perspective taking		0.132***	0.129***	0.078***
		(0.002)	(0.002)	(0.002)
T1 Personal traits				-0.023***

				(0.001)
T2 Family				0.048***
				(0.001)
T3 Openness				0.054***
				(0.001)
T4 Hobby				-0.004**
				(0.001)
T5 Work				-0.041***
				(0.001)
T6 Local expert				0.039***
				(0.001)
T7 Accommodation				0.077***
				(0.001)
T8 Service				0.062***
				(0.002)
T9 Origin				-0.024***
				(0.002)
T10 Travel				0.057***
				(0.001)
Constant	-0.286***	-0.283***	-0.283***	-0.304***
	(0.003)	(0.002)	(0.002)	(0.02)
R^2	0.0435	0.1350	0.1462	0.2559
Adjust R^2	0.0435	0.1349	0.1461	0.2557

	VIF				
range		1.0-1.21	1.0-1.21	1.0-1.24	1.03-1.70
	Observ	73,654	73,654	73,654	73,654
ations					

Notes: coefficients are shown in the table; standard errors are shown in parentheses; *p $<$ 0.1, **p $<$ 0.05, ***p $<$ 0.01.

4.1.1. *The effect of linguistic features*

The results show that the readability of self-descriptions is positively associated with the trust perceptions of guests (β = 0.007, p $<$ 0.01). For guests, the readability of self-descriptions can affect the process of acquiring information, and higher readability facilitates understanding the semantic information contained in self-descriptions. Consequently, higher readability can be helpful in improving the persuasive effect and further help to earn the trust of guests.

The relationship between sentiment intensity expressed in self-description and trust perception was also examined. The coefficients of both the linear and quadratic terms were significant (linear, β = 0.032, p $<$ 0.01; quadratic, β = -0.052, p $<$ 0.01). This evidence suggests that the relationship between sentiment intensity and trust perception can be described by an inverted U-shaped function. More positive sentiment being expressed in self-description is related to higher trust perception, but only up to a point. When the sentiment intensity is beyond the point, the trust perception will be reduced.

The beta coefficient of perspective taking was also positive and significant (β = 0.078, p $<$ 0.01). This finding suggests that employing second-person pronouns in self-description can improve trust perception.

4.1.2. *The effect of semantic features*

Model-4 examined the influence of semantic topics on trust perception. It can be observed that all semantic topics were significantly correlated with trust perception. Specifically, the topics of T2 "Family", T3 "Openness", T6 "Local expert", T7 "Accommodation", T8 "Service" and T10 "Travel" have a positive influence on trust perception. By contrast, the topics of T1 "Personal traits", T4 "Hobby", T5 "Work" and T9 "Origin" have a negative influence on trust perception. The result suggests that different semantic topics expressed in self-description play different roles in trust building:

(1) On the one hand, the hosts describing themselves in aspects of openness, enthusiasm, hospitableness and competence in providing better service are likely to obtain higher trust perception. On the other hand, hosts focusing on their age, occupation and personal interests when presenting themselves can only obtain lower trust perception. The reason for this could be that guests care more about the social ability and service ability of hosts. Open, warm, hospitable and competent hosts are more trustworthy for guests. In contrast, age, occupation and personal interests do not matter much in trust building.

(2) According to social penetration theory [52], hosts who prefer to provide deep and broad information, such as details about their family, service attitude, emotion, cultures and ability, can obtain higher trust perception. This is because that they offer deeper layers of disclosure than those hosts who only disclose superficial information, i.e., general personal information such as age, occupation and hobby.

(3) From the perspective of social exchange [53], guests pay more attention to their

psychological needs of obtaining unique experiences, varieties, and social relations. The topics of T2 "Family", T3 "Openness", T6 "Local expert", T7 "Accommodation", T8 "Service" and T10 "Travel" can satisfactorily meet guests' needs, which can further help build trust.

Additionally, it can also be seen that the control variable of notional words has a significant positive influence on trust perception. This result is consistent with the work of Zhang [11] and Ma [12].

4.2. The effect of trust perception on purchase behavior

We assume that order quantity can be used to quantify purchase behavior. However, the Airbnb website does not display order quantity. Thus, researchers often convert review quantity to estimated order quantity through the review rate. By crawling and analyzing the Airbnb data from 2008 to 2016, Ke [54] estimated the review rate is 18.6%. The Budget and Legislative Analyst's Office used two values of 72% and 30.5% for review rate in their report[6]. Although review rate 30.5% is more fact based, it didn't consider missing reviews because of deleted listings. The analysis of Inside Airbnb suggests that it would be reasonable to use 50% as the review rate because it sits almost exactly between 72% and 30.5%[7]. Accordingly, we used 50% as the review rate. We also found that review quantity is roughly linearly related with online sales and can be considered a valid proxy variable for online sales [55]. Thus, we estimated order quantity with the accumulated number of reviews received by a host. It should be noted that different values of review rate are equivalent for the following regression analysis. To ensure that the accommodations were linked with actual transactions, we only reserved listings with at least one guest review and the size of listings was reduced to 57,683.

Table 5: Descriptions and measurements of control variables

Variables	Descriptions and measurements
Superhost	A host with the superhost badge is labeled 1 and 0 otherwise.
Host since	The length of the time since a host registered on Airbnb (#days).
Verified status	Verified host =1, non-verified host = 0.
Response time	Within an hour=1, a few hours=2, hours=3, a few days=4, days=5.
Response rate	The rate of one host responding to his guests.
Price	The price of the listing (unit: dollars).
Listing type	Shared room=0, private room=1 and entire home=2.
Cancellation	Cancellation policies: flexible=0, moderate=1 and strict=2.
Accommodates	The number of guests accommodated by a given listing.
City_HongKong	Dummy variable: 1 if the city is Hong Kong and 0 otherwise.
City_NewYork	Dummy variable: 1 if the city is New York and 0 otherwise.
City_Paris	Dummy variable: 1 if the city is Paris and 0 otherwise.

Nine variables related to host reputation, response behavior and listing attributes were controlled for, as they could influence purchase behavior [11]. In addition, we controlled for the

---

[6] https://sfbos.org/sites/default/files/FileCenter/Documents/52601-BLA.ShortTermRentals.051315.pdf
[7] http://insideairbnb.com/about.html

city variable. The description and measurements of control variables are shown in Table 5. As shown in Table 6, the VIF values suggest that the collinearity between these independent variables does not warrant concern.

Table 6: Results of the negative binomial regression model

Variables	Dependent variable: purchase behavior	
	Model-5	Model-6
Control variables		
Superhost	0.366***(0.005)	0.353***(0.005)
Host since	0.221***(0.006)	0.219***(0.005)
Verified status	0.075***(0.005)	0.071***(0.005)
Response time	-0.020***(0.006)	-0.021***(0.006)
Response rate	0.068***(0.006)	0.066***(0.006)
Price	-0.024***(0.007)	-0.018**(0.007)
Listing type	-0.193***(0.007)	-0.183***(0.007)
Cancellation	0.309***(0.005)	0.299***(0.005)
Accommodates	0.158***(0.006)	0.142***(0.006)
City_HongKong	0.104***(0.017)	0.100***(0.017)
City_NewYork	0.186***(0.012)	0.212***(0.012)
City_Paris	0.287***(0.016)	0.322***(0.016)
Independent variables		
Trust perception		0.140***(0.005)
Constant	3.006***(0.008)	2.983***(0.008)
Log likelihood	-237,925.97	-237,522.97
LR X^2	14,466.06	15,272.07
Pseudo R^2	0.0295	0.0311
VIF range	1.04-1.97	1.05-1.97
Observations	57,683	57,683

Notes: coefficients are shown in the table; standard errors are shown in parentheses; *p $<$ 0.1, **p $<$ 0.05, ***p $<$ 0.01.

Negative binomial regression was used to identify the effect of trust perception on purchase behavior. Table 6 shows the regression results. Model-5 presents the effects of control variables, and Model-6 includes the effect of trust perception on purchase behavior. The beta coefficient of trust perception was positive and significant (β = 0.140, p $<$ 0.01), and the result suggests that the trust perception of guests regarding host self-description can promote purchases on Airbnb.

Moreover, all control variables were significantly correlated with purchase behavior. The superhost badge, host since, verified status, quick response and high response rate are positively associated with purchase behavior. These results are consistent with Zhang [11]. Listings offering an entire room receive more purchases than those for private and shared rooms. Higher prices will reduce purchases, which follows market rules. These results are in agreement with Wu [10]. We also found that a strict cancellation policy is helpful.

It should be noted that the regression analysis was performed on four cities' active listings during one year, which suggests the reliability of the conclusions. We also found that the trust

perception of self-descriptions in London is significantly higher than that in Paris (see Table 4), and the review quantity of hosts in London is less than the other three cities (see Table 5). However, we did not delve further into the phenomena because this is not the focus of this work.

5. Discussion and conclusion

Trust has been widely considered an urgent problem on Airbnb. However, how the self-description of hosts can influence guests' trust perception and whether trust perception can lead to purchase behavior remain unclear. This paper proposed a text analytics framework for understanding the relationships among self-description, trust perception and purchase behavior. In detail, we proposed a deep-learning procedure to automatically code guests' trust perception. Additionally, the linguistic and semantic features were extracted from self-description using text mining methods. A series of regression analyses were conducted to determine the impact of linguistic and semantic features on trust perception. Finally, the relationship between trust perception and purchase behavior was investigated. Several interesting findings of this work can help understand the mechanism of trust-building and purchase decisions on Airbnb.

5.1. Research findings

Many studies have investigated the influence of the amount of information contained in self-description on trust perception [11, 12]. However, the influence of linguistic and semantic features of self-description on trust perception remains under-researched. In this paper, we investigated the full spectrum of textual features influencing trust perception through quantitative analysis on Airbnb. We found that readable self-description can help build trust. The complexity of words and sentences used in self-description may weaken the persuasive effect. With regard to sentiment, we found that hosts describing themselves with excessive positive sentiment tend to receive lower trust perceptions, and this finding differs from those of previous studies [11, 12, 37, 38]. Moreover, hosts showing the ability of perspective taking in their self-descriptions can obtain higher trust. The finding suggests that hosts should not only describe themselves in detail but also need to pay more attention to the needs of guests.

Regarding semantic content, the existing literature mainly relies on human coding or word clustering to identify topics from self-descriptions [3, 12, 26]. However, human coding is not applicable to large-scale data, and it is difficult to choose the best number of topics. In this paper, we identified the topics of self-descriptions using LDA and by refering to topic coherence [49], we could obtain relatively comprehensive topics from self-descriptions. By analysing the relationship between the identified topics and trust perception, we found that hosts who tend to make profound disclosures and provide deep and broad information, such as details about their family members, openness, service and travel experiences, can obtain higher trust perception. This disclosed information can satisfactorily meet guests' psychological needs of experiencing, communication, and interaction. In contrast, hosts who mainly focus on their age, occupation and personal interests when presenting themselves can only obtain lower trust perception. Although this information can help guests understand hosts, guests do not feel the enthusiasm and warmth of hosts. These findings extend the work of [3, 11, 12].

Finally, we found that guests' trust perceptions towards hosts' self-descriptions can promote purchases. To the best of our knowledge, we are the first to examine the relationship between trust perception and purchase behavior on Airbnb. We observed that the linguistic and semantic features of self-description can impact trust perception, and trust perception can further

impact purchase behavior. Thus, the importance of making proper self-descriptions for hosts on Airbnb should be emphasized.

5.2. *Theoretical and practical contributions*

From a theoretical perspective, existing studies have found that response to feedback, verifications, profile photos, and "superhost" status are significantly correlated with trust [4, 6, 8, 9, 10]. However, few studies have focused on self-descriptions of hosts, despite it being an important channel for guests to gain deeper and comprehensive understandings of hosts. We knew little about what features of hosts' self-descriptions can influence guests' trust perceptions and how trust perceptions can further influence their purchase decisions. Our study fills this gap. We conducted a comprehensive and in-depth analysis of the influence of linguistic and semantic content of self-description on perceived trust on Airbnb. In addition, we also examined the relationship between trust perception towards hosts' self-descriptions and guests' purchase behavior.

With regard to the methods of trust-related research, the majority of prior studies have conducted experiments or designed questionnaires to allow participants to code trust perception [3, 8, 12, 26]. These methods of experiments and surveys are usually time-consuming and limited to small-scale data. In this paper, we developed a deep learning procedure to automatically code a trust perception score for each self-description text based on AB-BiLSTM, which can be readily applied to a large amount of data. Our approach provides an alternative to survey-based or human coding-based perceived trust measurement.

Regarding the sentiment of self-description, positive sentiment expression was found to be positively related to guests' trust [11]. However, sentiment intensity is rarely studied in the context of the sharing economy. Our work found that either too low or too high positive sentiment intensity expressed in self-description is harmful for trust perception. This finding may direct future studies to pay more attention to the negative side of sentiment expression in the sharing economy. In addition, we are the first to find that hosts who adopt perspective taking tend to be perceived as more trustworthy.

Moreover, we also found that the readability of host description is positively related to trust perception. The analysis of the topics of self-description provided useful instructions for hosts on the choice of semantic content, which serves as a supplement for existing studies [3, 11, 12].

Finally, the analysis is based on large-scale real-world data of Airbnb hosts across different cultures. Text mining technologies, i.e., tokenization, sentiment analysis, and topic models, were applied to extract and measure textual features, which can be readily applied to a large amount of data. This is a supplement to existing studies that mainly relied on human coding, surveys and experimental approaches [8, 9, 12, 26]. The methods used in our research also enrich the application of text mining methods in trust building for short-term rentals.

From a practical perspective, our findings can help hosts improve self-descriptions, which can further increase their perceived trustworthiness and boost order quantity. As hosts will attempt to build trust with guests by presenting themselves as trustworthy, we assert that hosts should consciously manage their presentation. The findings of this paper can help service providers better manage their self-descriptions. For example, when hosts write their self-descriptions, they are advised to avoid using complex words and sentences, which can increase the difficulty of reading and reduce the persuasive effect of the self-descriptions. Instead, they should use more notional words (e.g., nouns, adjectives, verbs) that can provide more context and detailed information about hosts and further reduce the uncertainty of guests. More important, hosts should carefully and

wisely use emotional words and second-person pronouns when they describe themselves. Additionally, hosts should not only disclose their demographic-related information, such as age, occupation and origin, but also focus on expressing their family relationships, openness and eagerness to meet new people, attitudes towards service, service ability and travel experience, which are more important for meeting guests' needs. The findings in this paper not only benefit the improvement of short-term rental sites but can also provide guidance for other sharing economy platforms. It is advisable that sharing platform managers or designers pay more attention to the self-disclosure mechanism, especially regarding self-description.

5.3. *Limitations and directions for future research*

This paper broadens the research concerning trust building in the sharing economy and enriches the available research methods by introducing text mining technology. The findings can help hosts develop better self-descriptions and can also provide instructions for building more user-friendly sharing platforms. However, this study has several limitations. First, we used data from a single short-term rental site—Airbnb. Thus, it would be interesting to extend the research to other kinds of sharing services. Second, in order to dilute the effect of cultural background on trust perception and accommodation selection, we selected four cities including New York, London, Paris and Hong Kong to ensure a diverse sample of hosts and guests, but we did not investigate the reason behind the difference of these cities on trust perception and purchase behavior, which is perhaps an interesting research direction.

Acknowledgments

References

[1]   A. G. Mauri, R. Minazzi, M. Nieto-García, G. Viglia, Humanize your business. the role of personal reputation in the sharing economy, International Journal of Hospitality Management 73 (2018) 36–43.   doi:10.1016/j.ijhm.2018.01.017 .

[2]   D. Guttentag, Airbnb: disruptive innovation and the rise of an informal tourism accommodation sector, Current Issues in Tourism 18 (12) (2015) 1192–1217. doi:10.1080/13683500.2013.827159 .

[3]   I. P. Tussyadiah, S. Park, When guests trust hosts for their words: Host description and trust in sharing economy, Tourism Management 67 (2018) 261 – 272. doi:10.1016/j.tourman.2018.02.002 .

[4]   J. Kim, Y. Yoon, H. Zo, Why people participate in the sharing economy: A social exchange perspective, in: Proceedings of the 2015 pacific Asia conference on information

systems, 2015, pp. 1–6.

[5]   A. Wilson, V. A. Zeithaml, M. J. Bitner, D. D. Gremler, Services marketing: Integrating customer focus across the firm, McGraw Hill, 2012.

[6]   S. Liang, M. Schuckert, R. Law, C. C. Chen, Be a "superhost": The importance of badge systems for peer-to-peer rental accommodations, Tourism Management 60 (2017) 454–465. doi:10.1016/j.tourman.2017.01.007 .

[7]   G. Zervas, D. Proserpio, J. Byers, A first look at online reputation on airbnb, where every stay is above average, available at SSRN: https://ssrn.com/abstract=2554500 (accessed 28 January 2015) (2015) 1–22.

[8]   E. Ert, A. Fleischer, N. Magen, Trust and reputation in the sharing economy: The role of personal photos in airbnb, Tourism Management 55 (Supplement C) (2016) 62–73. doi:10.1016/j.tourman.2016.01.013 .

[9]   A. Fagerstrøm, S. Pawar, V. Sigurdsson, G. R. Foxall, M. Y. de Soriano, That personal profile image might jeopardize your rental opportunity! on the relative impact of the seller's facial expressions upon buying behavior on airbnb, Computers in Human Behavior 72 (Supplement C) (2017) 123–131.   doi:10.1016/j.chb.2017.02.029 .

[10]   J. Wu, P. Ma, K. Xie, In sharing economy we trust: The effects of host attributes on short-term rental purchases, International Journal of Contemporary Hospitality Management 29 (11) (2017) 2962–2976.   doi:10.1108/IJCHM-08-2016-0480 .

[11]   L. Zhang, Q. Yan, L. Zhang, A computational framework for understanding antecedents of guests' perceived trust towards hosts on airbnb, Decision Support Systems 115 (2018) 105 – 116.   doi:10.1016/j.dss.2018.10.002 .

[12]   X. Ma, J. T. Hancock, K. Lim Mingjie, M. Naaman, Self-disclosure and perceived trustworthiness of airbnb host profiles, in: Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing, 2017, pp. 2397–2409. doi:10.1145/2998181.2998269 .

[13]   D. M. Blei, A. Y. Ng, M. I. Jordan, Latent dirichlet allocation, Journal of machine Learning research 3 (Jan) (2003) 993–1022.

[14]   R. Botsman, R. Rogers, What's mine is yours: how collaborative consumption is changing the way we live, Collins, 2011.

[15]   J. Barbalet, Self-interest and the theory of action., British Journal of Sociology 63 (3) (2012) 412–429.   doi:10.1111/j.1468-4446.2012.01417.x .

[16]   A. Jøsang, R. Ismail, C. Boyd, A survey of trust and reputation systems for online service provision, Decision Support Systems 43 (2) (2007) 618–644. doi:10.1016/j.dss.2005.05.019 .

[17]   J. Bridges, C. Vásquez, If nearly all airbnb reviews are positive, does that make them meaningless?, Current Issues in Tourism 21 (18) (2018) 2057–2075. doi:10.1080/13683500.2016.1267113 .

[18]   A. Beldad, M. de Jong, M. Steehouder, How shall i trust the faceless and the intangible? a literature review on the antecedents of online trust, Computers in Human Behavior 26 (5) (2010) 857 – 869.   doi:10.1016/j.chb.2010.03.013 .

[19]   C. R. Berger, R. J. Calabrese, Some Explorations in Initial Interaction and Beyond: Toward a Developmental Theory of Interpersonal Communication, Human Communication Research 1 (2) (2006) 99–112.   doi:10.1111/j.1468-2958.1975.tb00258.x .

[20]   J. Delhey, K. Newton, C. Welzel, How general is trust in "most people"? solving the radius of trust problem, American Sociological Review 76 (5) (2011) 786–807.

[21]   D. Christozov, S. Chukova, P. Mateev, A measure of risk caused by information asymmetry in e-commerce., Issues in Informing Science & Information Technology 3 (2006) 147 – 158.   doi:10.28945/879 .

[22]   A. Usman, S. Okafor, Exploring the relationship between social media and social influence, in: Leveraging Computer-Mediated Marketing Environments, IGI Global, 2019, pp. 83–103.   doi:10.4018/978-1-5225-7344-9.ch004 .

[23]   C. Cheshire, Online trust, trustworthiness, or assurance?, Daedalus 140 (4) (2011) 49–58.   doi:10.1162/DAED_a_00114 .

[24]   P. Chandon, V. G. Morwitz, W. J. Reinartz, Do intentions really predict behavior? self-generated validity effects in survey research, Journal of Marketing 69 (2) (2005) 1–14. doi:10.1509/jmkg.69.2.1.60755 .

[25]   E. Kytö, M. Virtanen, S. Mustonen, From intention to action: Predicting purchase behavior with consumers? product expectations and perceptions, and their individual properties, Food Quality and Preference 75 (2019) 1–9.   doi:10.1016/j.foodqual.2019.02.002 .

[26]   X. Ma, T. Neeraj, M. Naaman, A computational approach to perceived trustworthiness of airbnb host profiles, in: ICWSM, 2017, pp. 604–607.

[27]   S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.   doi:10.1162/neco.1997.9.8.1735 .

[28]   A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, in: Proceedings of the 2017 conference on neural information processing systems, 2017, pp. 5998–6008.

[29]   P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, B. Xu, Attention-based bidirectional long short-term memory networks for relation classification, in: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 2016, pp. 207–212. doi:10.18653/v1/P16-2034 .

[30]   R. C. Mayer, J. H. Davis, F. D. Schoorman, R. C. Mayer, J. H. Davis, An integrative model of organizational trust, Academy of management review 20 (3) (1995) 709–734.

[31]   J. Pennington, R. Socher, C. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 conference on empirical methods in natural language processing, 2014, pp. 1532–1543.   doi:10.3115/v1/D14-1162 .

[32]   C. L. Toma, J. T. Hancock, What lies beneath: The linguistic traces of deception in online dating profiles, Journal of Communication 62 (1) (2012) 78–97. doi:10.1111/j.1460-2466.2011.01619.x .

[33]   L. Larrimore, L. Jiang, J. Larrimore, D. Markowitz, S. Gorski, Peer to peer lending:

The relationship between language features, trustworthiness, and persuasion success, Journal of Applied Communication Research 39 (1) (2011) 19–37.

[34]   N. Hu, I. Bose, N. S. Koh, L. Liu, Manipulation of online reviews: An analysis of ratings, readability, and sentiments, Decision support systems 52 (3) (2012) 674–684. doi:10.1016/j.dss.2011.11.002 .

[35]   B. L. Zakaluk, S. J. Samuels, Readability: Its Past, Present, and Future, Intl Literacy Assn, 1988.

[36]   R. Flesch, How to write plain English: A book for lawyers and consumers, Harper & Row New York, NY, 1979.

[37]   S. M. Norman, B. J. Avolio, F. Luthans, The impact of positivity and transparency on trust in leaders and their perceived effectiveness, The Leadership Quarterly 21 (3) (2010) 350–364.   doi:10.1016/j.leaqua.2010.03.002 .

[38]   S. D. Pugh, Service with a smile: Emotional contagion in the service encounter, Academy of management journal 44 (5) (2001) 1018–1027.

[39]   D. Yin, S. D. Bond, H. Zhang, Keep your cool or let it out: Nonlinear effects of expressed arousal on perceptions of consumer reviews, Journal of Marketing Research 54 (3) (2017) 447–463.   doi:10.1509/jmr.13.0379 .

[40]   Z. Wang, H. Mao, Y. J. Li, F. Liu, Smile big or not? effects of smile intensity on perceptions of warmth and competence, Journal of Consumer Research 43 (5) (2017) 787–805. doi:10.1093/jcr/ucw062 .

[41]   M. L. Jensen, J. M. Averbeck, Z. Zhang, K. B. Wright, Credibility of anonymous online product reviews: A language expectancy perspective, Journal of Management Information Systems 30 (1) (2013) 293–324.   doi:10.2753/MIS0742-1222300109 .

[42]   R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, C. Potts, Recursive deep models for semantic compositionality over a sentiment treebank, in: Proceedings of the 2013 conference on empirical methods in natural language processing, 2013, pp. 1631–1642.

[43] A. Gerace, A. Day, S. Casey, P. Mohr, An exploratory investigation of the process of perspective taking in interpersonal situations, Journal of Relationships Research 4 (2013) 1–12. doi:10.1017/jrr.2013.6 .

[44] R. S. Marvin, M. T. Greenberg, D. G. Mossler, The early development of conceptual perspective taking: Distinguishing among multiple perspectives, Child Development 47 (2) (1976) 511–514. doi:10.2307/1128810 .

[45] A. R. Todd, G. V. Bodenhausen, J. A. Richeson, A. D. Galinsky, Perspective taking combats automatic expressions of racial bias, Journal of personality and social psychology 100 (6) (2011) 1027–1042. doi:10.1037/a0022308 .

[46] A. D. Galinsky, W. W. Maddux, D. Gilin, J. B. White, Why it pays to get inside the head of your opponent: The differential effects of perspective taking and empathy in negotiations, Psychological science 19 (4) (2008) 378–384. doi:10.1111/j.1467-9280.2008.02096.x .

[47] C.-H. Peng, D. Yin, C.-P. Wei, H. Zhang, Impact of perspective taking on reviewer behavior: A multi-method exploration, in: Proceedings of the 2017 international conference on information systems, 2017, pp. 1–16.

[48] J. W. Pennebaker, M. R. Mehl, K. G. Niederhoffer, Psychological aspects of natural language use: Our words, our selves, Annual review of psychology 54 (1) (2003) 547–577. doi:10.1146/annurev.psych.54.101601.145041 .

[49] D. Newman, J. H. Lau, K. Grieser, T. Baldwin, Automatic evaluation of topic coherence, in: Proceedings of the 2010 annual conference of the North American chapter of the ACL, 2010, pp. 100–108.

[50] P. J. Rousseeuw, A. M. Leroy, Robust regression and outlier detection, Vol. 589, John wiley & sons, 2005.

[51] R. M. O'brien, A caution regarding rules of thumb for variance inflation factors, Quality & quantity 41 (5) (2007) 673–690. doi:10.1007/s11135-006-9018-6 .

[52] I. Altman, D. A. Taylor, Social penetration: The development of interpersonal relationships, New York, Holt, Rinehart and Winston, 1973.

[53] C. J. Lambe, C. M. Wittmann, R. E. Spekman, Social exchange theory and research on business-to-business relational exchange, Journal of Business-to-Business Marketing 8 (3) (2001) 1–36. doi:10.1300/J033v08n03_01 .

[54] Q. Ke, Sharing means renting?: An entire-marketplace analysis of airbnb, in: Proceedings of the 2017 ACM on Web Science Conference, ACM, 2017, pp. 131–139. doi:10.1145/3091478.3091504 .

[55] Q. Ye, R. Law, B. Gu, The impact of online user reviews on hotel room sales, International Journal of Hospitality Management 28 (1) (2009) 180 – 182. doi:10.1016/j.ijhm.2008.06.011 .
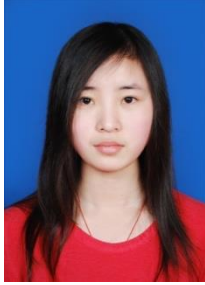
Le Zhang is a Ph.D. candidate in the School of Economics and Management, Beijing University of Posts and Telecommunications, Beijing, China. Her research interests include trust in the sharing economy, social computing, natural language processing.

Qiang Yan is a professor in the School of Economics and Management at Beijing University of Posts and Telecommunications. His researches focus on E-commerce and information systems. He has published more than 100 papers in peer-reviewed journals.

Leihan Zhang Computer research natural
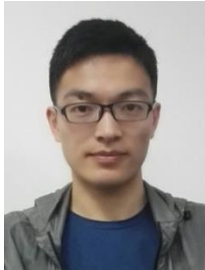
is a research associate in the Wangxuan Institute of Technology, Peking University, Beijing, China. His interests include data mining, complex network, and language processing.

Credit Author Statement

Le Zhang: Conceptualization, Methodology, Software, Writing-Original Draft, Validation, Investigation, Resources.

Qiang Yan: Conceptualization, Supervision, Project administration, Funding acquisition.

Leihan Zhang: Formal analysis, Data curation, Methodology, Software, Writing- Review and Editing, Visualization.

Highlights:

(1) A deep-learning procedure is proposed to predict guests' trust perception.
(2) The readability and perspective taking in self-description can help build trust.
(3) Excessive positive sentiment expression would decrease trust perception.
(4) More introduction on family, openness, service and travel in self-description would be helpful.
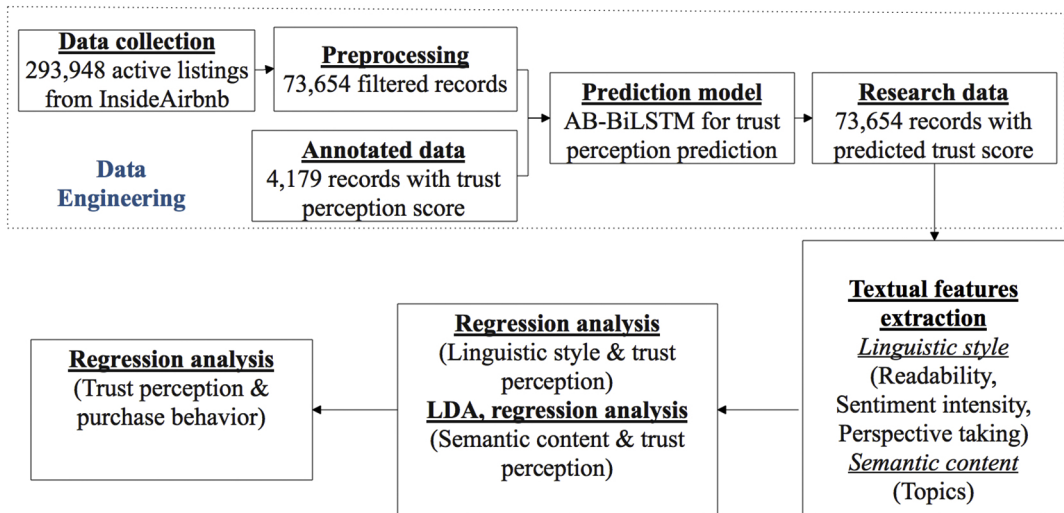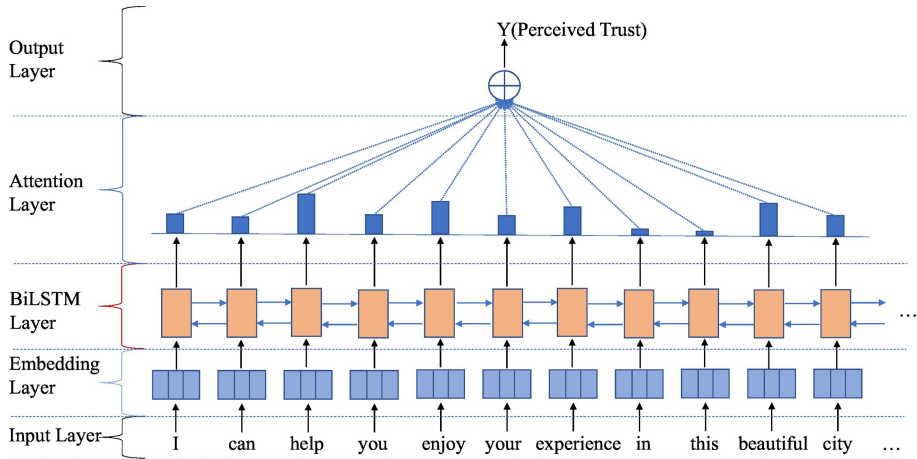(5) The perceived trust derived from self-description does contribute to purchases.

Figure 1

Figure 2