

Exploring Trust in Online Ride-sharing Platform in China: A Perspective of Time and Location

Xiangsong Meng
University of
International Business
and Economics
xiangsong_meng@163.com

Xusen Cheng
University of
International Business
and Economics
xusen.cheng@uibe.edu.cn

Shixuan Fu
Beijing International
Studies University
fsx8888@163.com

Jianshan Sun
Hefei University of
Technology
sunjs9413@hfut.edu.cn

Abstract

Trust is a key issue to be considered deliberately in the online ride-sharing platform to reduce risk and ensure transactions. In this paper, trust-in-platform is explored from these two perspectives to fill the research gaps. A ride-sharing platform in China was investigated. Results show that trust-in-platform in economically developing districts is slightly higher than that in economically developed districts. At the same time, trust-in-platform level differs in time, trust-in-platform levels are obviously lower between 19'o clock and 23'o clock. Moreover, machine learning is employed to predict the relationships between time/location and trust-in-platform. The result is that recall is 78.3%, precision is 57.3%, and F1 is 66.2%. The result shows trust-in-platform has an obvious correlation with time and location, thus further consolidates the findings. This study contributes to the existing knowledge on trust in the ride-sharing platforms and has practical implications for platform operators.

1. Introduction

With the rapid growth of information technology, the past decade has witnessed the thriving of digital sharing economy. The sharing economy provides services in the form of renting, which used to have access to services in the form of buying [13]. The logic behind sharing economy is that users should focus on the right to use an asset rather than ownership, which encourages individuals to share his/her property or services with others without ownership transfer [1]. Sharing economy facilitates instant renting business and helps individuals change fixed mindset related to sale and purchase [2]. Individuals have benefited a lot

from the sharing economy, including rented cars (Uber), rented bicycles (OfO), and even rented bedrooms when travelling (AirBnB).

It has been well acknowledged that trust facilitates transactions under the circumstances of risk, uncertainty and interdependence [3]. In contrast to traditional economy based on corporate reputation, sharing economy greatly depends on peer to peer communication [4], thus is easy to be imposed with risks and uncertainties. For example, prior to a ride-sharing arriving, the passenger may worry about being late for work and thus make an alternative choice that takes a taxi in his eyes, the driver may cancel the order when the destination is so remote that it is possible to return empty. However, trust serves as a bond to linking strange passengers and drivers [5]. Once building trust, peer to peer transactions will increase the ability to resist risks [5].

Trust refers to the positive expectations with regard to the conduct, motives, and intentions of trustees [23]. In terms of the ride-sharing platform context, trust-in-platform is a willingness to complete a transaction through the platforms, and requires driver and passenger to work together to get the transaction done. Existing studies have investigated trust from the perspective of individual personality differences [6], platform design [7], and user generated contents [8]. However, the relationship between time/location and trust-in-platform has been scarcely investigated. Rush time and heavy traffic often lead to long waiting time that influence users' trust perception and further change individuals' use intention.

Many studies have been carried out through survey [4], interview [9] and experiment [10], but these three methods are not easy to cope with massive information. In previous studies, using machine learning approach can solve this problem well. Cheng et al. [8] used deep learning to explore factors influencing on trust perception depending on user review data. Liu et al. [20] used decision tree algorithm to predict the trust

levels based on personal historical feature. Moreover, considering that previous studies make use of specific individual's behavior to derive an individual's trust level [11][12], machine learning based on all the historical data is capable of processing massive information simultaneously and get great performance.

In this study, two questions will be addressed: (1) whether time and location have a positive correlation with trust in the online ride-sharing platform? (2) if so, what are the relationships?

2. Research background

2.1. Sharing economy and ride-sharing

The sharing economy, a peer-to-peer economy, has access to services through renting rather than buying [13]. The ride-sharing is one important part of sharing economy. There are two distinct characteristics: one-time-only deals and a strange driver/passenger. Since one-time-only deals involve a lot of financial risks and strangers maybe cause security problems, successful business requires trust to develop reputation mechanisms [14]. Kim et al. [15] considered trust as a premise of consumer decisions on online transaction. Consequently, ride-sharing platform is difficult to operate continuously without trust. A trust failure may not only frustrate the deal but also jeopardize the integrity of a well-developed platform [16].

Many researchers conduct study related to trust in the ride-sharing. Cheng et al. [17] investigated which factors can influence online and offline service quality with respect to ride-sharing. Mazzella [18] evaluated the levels of trust in different familiarity groups to verify ride-sharing available among strangers. However, the relationship between time/location and trust-in-platform is supposed to attract some attention. Rush time and heavy traffic often lead to so long waiting time that influence users' trust perception in the context of ride-sharing. In such basis, ride-sharing scholars and platform operators are supposed to attach more importance to trust.

2.2. Trust

McKnight et al. [19] defined trust as "one believes in and are willing to depend on, another party". Based on this point, trust is defined as positive expectations with regard to the conduct, motives, and intentions of trustees [23], it causes a willingness to complete a transaction requires driver and passenger to work together in the context of the ride-sharing.

Fang et al. [22] used online survey method to verify that trust have a considerable and positive effect on

consumers' intention to repurchase. Consequently, trust plays an important role in the sharing economy platform.

Many researchers have investigated trust from different perspectives. Gefen et al. [6] found that consumer trust is the same importance as perceived usefulness and ease of use in the context of online shopping. Ghose et al. [7] found that consumer behavior on social media and search engines is closely related. Although researchers have considered trust from multiple dimensions, in this paper, we explored trust from these two perspectives of time and location in the ride-sharing platform.

What's more, most studies have been carried out through survey [4], interview [9] and experiment [10]. However, today has entered the information age and big data is more and more valuable. Hence, data mining widely used in various fields at present. Cheng et al. [8] used deep learning to explore factors influencing on trust perception depending on user review data. Liu et al. [20] used decision tree to predict the trust levels based on agents' specific feature with respect to online auction. Obviously, machine learning based on all the historical data is capable of processing massive information simultaneously and get better performance. Consequently, our proposal employs machine learning to estimate the trust-in-platform levels of a new order on the basis of time and location.

3. Research method

We employed traditional statistical analysis and machine learning approach in our research.

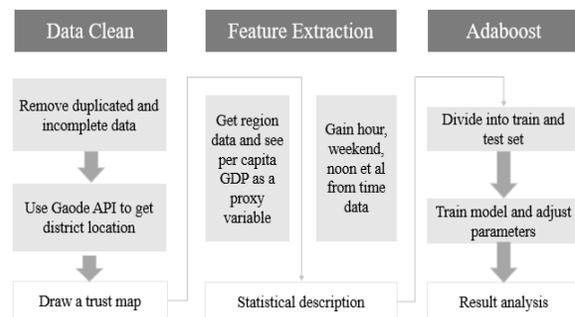


Figure 1. Research Framework

Figure 1 illustrates the research framework: Firstly, we carried out data cleaning, including deleting duplicated data and data with missing key values, then Gaode API were applied to get district location; Secondly, we drew a trust map through the amount of low trust-in-platform orders; Thirdly, we constructed time and location variables and conducted statistics

analysis; Finally, we conducted experiments based on SVM (Support Vector Machine), LR (Logistic Regression) and AdaBoost model to verify the relationships.

4. Data cleaning

The raw data includes order number, pickup time, pickup longitude, pickup latitude, terminal longitude, terminal latitude, total fee. The variable named “total fee” means that non-zero value shows successful matched order and zero values indicates that the order was unmatched. Based on the previous definition of trust-in-platform, we set matched order as a high trust signal, and unmatched order as a low trust signal.

The data from an online ride-sharing platform covers one month in Beijing. Before data analysis, there is a must to introduce data processing tool. Python is an object-oriented scripting language. Pandas based on python (Python Data Analysis Library) is an open source library with high-performance data analysis tools.

After removing the duplicated, incomplete and abnormal data from all raw data, there are still 43598 record leaving. Removing invalid data ensures the validity of the experimental results and sufficient data volume guarantees the stability of the next experiment, which helps us analyze the correlation between trust-in-platform and time/location.

5. Statistical description

We set that matched order means high trust-in-platform, unmatched order means low trust-in-platform in this paper. Previous research has used the same mechanism for the proxy of trust-in-platform. For example, Liu et al. [20] viewed trust as a synonymous with successful transaction, predicting the trust level of a potential deal in the context of online auction.

5.1. Location dimension

In order to verify associations between location and trust-in-platform, we extracted concrete and quantitative data and compare the trust-in-platform levels in different regions according to the economic development of different regions.

Figure 2, a heat map, illustrates a trust map based on the distribution of ending points of low trust-in-platform orders during a day for city of Beijing, China, at scale of 30 km. The red area is a large concentration of low trust-in-platform level orders. The yellow area is a small distribution of orders with low trust-in-platform level. The green area is the boundary of order

distribution. We can see that the amount of low trust-in-platform orders in the middle area are more than that in the edged area. In addition, it is common sense that economic development in the Central Area of Beijing is better than that in other area. However, to verify that there is a correlation between trust-in-platform and location, it is necessary to calculate low trust-in-platform ratio in different districts. We employed the four steps to calculate low trust-in-platform ratio:

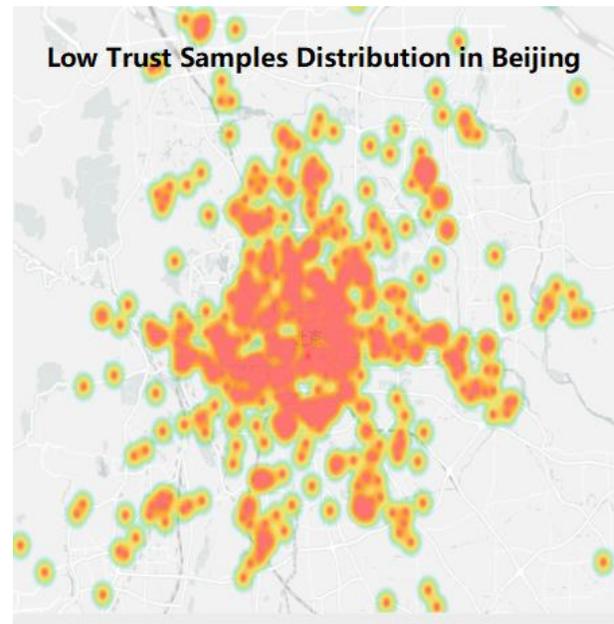


Figure 2. Trust Map

Step one: We employed Gaode Map API to get district data with processing the raw location data. Gaode Map API (<https://lbs.amap.com/>) is a leading LBS (Location Based Service) provider in China, with advanced data fusion technology and massive data processing capabilities. We finally generated administrative district data by comparing longitude and latitude data with map on Gaode Map API.

Step two: We divided orders by trust-in-platform levels and districts. On the basis of the results of matched orders, we split all orders by high/low trust-in-platform and divided them into two groups. Further, we grouped these two groups by district into smaller groups.

Step three: We counted the amount of high trust-in-platform and low trust-in-platform orders in order to calculating unmatched rate by district.

Based on these three steps, we have found that in the central areas, which are normally economically developed areas, the trust-in-platform levels are lower than those developing areas that are normally distributed outside the central city.

Research finding 1: Under the conditions of this study, the trust-in-platform levels in economically developing regions are comparatively higher than those in economically developed regions.

5.2. Time dimension

In addition, time has been taken into consideration. We split all matched orders in each hour of a day in one month.

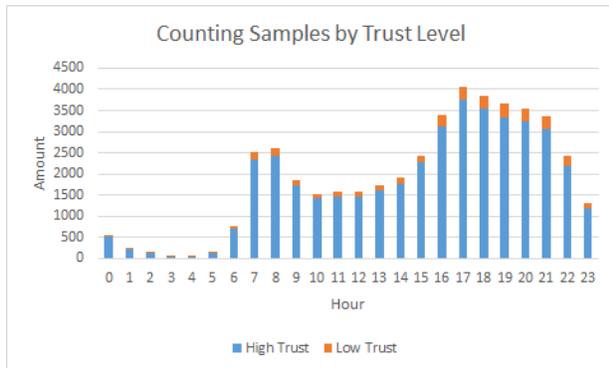


Figure 3. Trust Amount Counted by Time Period

Figure 3, the blue represents the number of high trust-in-platform orders while the orange represents that of low trust-in-platform orders. The X-axis refers to twenty-four hours in one day. The Y-axis refers to low trust-in-platform level samples amount. We ignored the datasets on 2'o clock, 3 o'clock and 4 o'clock since there were very few orders at these time.

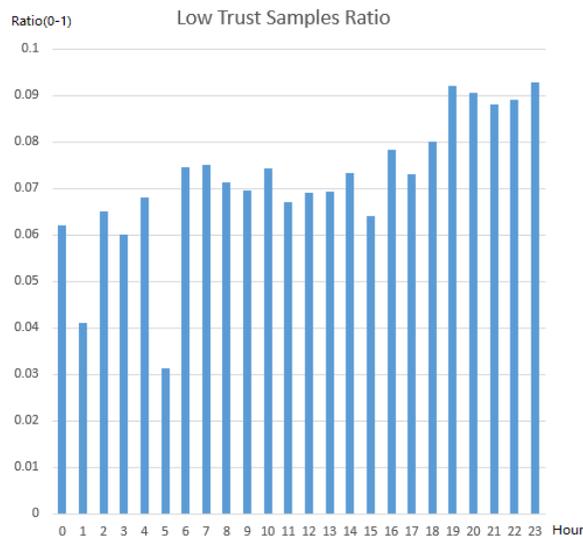


Figure 4. Trust Levels Categorized by Time Period

Figure 4, the bars represent ratio of unmatched orders in different time period of one day. The X-axis refers to twenty-four hours in one day. The Y-axis refers to the ratio of low trust-in-platform level samples. On the one hand, the trust-in-platform levels between 19'o clock and 23'o clock are obviously lower. The reason may be that some people are more worried about their safety at night. On another hand, the trust-in-platform levels are higher comparatively at 1 o'clock and 5 o'clock. We tend to think that the number of vehicles that can be selected during this time period is so small that people's willingness to cancel orders is reduced.

Research finding 2: Under the conditions of this study, the trust-in-platform levels between 19'o clock and 23'o clock are obviously lower, and the trust-in-platform levels are higher at 1 o'clock and 5 o'clock than those in other time periods.

6. Machine learning

Since we aim to predict trust-in-platform levels based on the factors of time and location, we should find some proxy variables for time and location. As for location, because location is not a numerical variable, location can't be directly brought into the model, thus it needs to be numerical first. Additionally, it was found that there was correlation between trust-in-platform level and per capita GDP. Therefore, the per capita GDP was used to represent differences between districts. As for time, we need to extract variables that are closely related to travel. The reason why we include the proxy variables is as follows. There is a must to merge time periods that belong to the same time periods. Thus, we divided time data into working hours, daytime, evening, morning peak, evening peak and weekend. The prediction model was conducted based on the above time periods.

In this paper, AdaBoost is used for establishing trust-in-platform model based on time and location factors and predicts the trust-in-platform levels of /unknown samples. In addition, to confirm the validity of the model, we view LR (logistic Regression) and SVM (Support Vector Machine) as benchmark. AdaBoost (Adaptive Boosting), a popular machine learning method, is an iterative algorithm [21] that transforms weak classifiers into strong classifiers. In every iteration, each sample classified incorrectly will be given larger weight and is expected to be identified and classified correctly in future. In this way, we had larger amount of correctly classified samples and then had a more effective model. We predict trust-in-platform levels by four steps:

Step 1: We first standardized the cleaned data in order not to change the distribution of the original data and reduce the effects of noise.

Step 2: We divided all samples into training set (67% of all samples) and test set (the remaining 33%). Train set is used to train model; test set is used to predict the results of new samples.

Step 3: We applied decision tree model to AdaBoost as a base classifier, with max_depth set to 2, min_samples_split set to 20, min_samples_leaf set to 5, the number of iterations set to 200 and learning rate set to 0.2. In our experiments, we took five sets of values for each parameter and then recorded the parameters of the highest f1 value. Finally, the local adjustment of the parameters was carried out to ensure that the Recall value is maximized under certain conditions of precision value.

Step 4: We used AdaBoost to conduct four experiments with LR, SVM as base classifiers based on these same data set, method and parameters. We compared these three method through recall value, precision value and f1 value.

Table 1 shows the prediction recall value, precision value and f1 value of test set in three experiments. Because our experimental goal is to maximize the recall rate while ensuring a certain accuracy, we found that Adaboost is superior to the first two algorithms in both recall and f1 values. The results show that on the one hand we can use the time and location factors to find out 78.3% of all low trust-in-platform orders in the future.; on the other hand, more than half of all the low trust-in-platform orders we forecast are correct. In addition, due to the size of the data set, the difference in runtime between the three is not obvious in the experiment.

Table 1. Test Dataset Predict Result

	SVM	LR	Adaboost
Recall	0.610	0.769	0.783
Precision	0.590	0.586	0.573
F1	0.600	0.659	0.662

Research finding 3: Based on the same test dataset, the effect of Adaboost method is significantly better than the other two methods when predicting trust-in-platform through time and location.

7. Conclusion

In this paper, we find that time and location have a correlation with trust in the ride-sharing platform. Based on the statistical descriptions, we found that there are indeed differences in terms of trust-in-

platform when considering different districts or time periods.

Firstly, as for location, the trust-in-platform in economically developing regions are comparatively higher than those in less developed regions.

Secondly, in terms of time, the trust-in-platform levels between 19'o clock and 23'o clock are obviously lower, and the trust-in-platform levels at 1 o'clock and 5 o'clock are obviously higher than other time.

Thirdly, in order to verify that these differences are meaningful for assess trust-in-platform, on the basis of the same control variables, we used machine learning to conduct three training and predicting experiments, including LR, SVM and Adaboost. Research results show that time and location are related to trust-in-platform. In addition, despite being in an early research stage, we find that the prediction f1 of result is 66.2%, which means that the relationship between time/location and trust-in-platform can't be ignored.

We found out the correlation between trust-in-platform and time/location, which few people mentioned before. The trust-in-platform levels in different time and location are significantly not the same. These two factors would be valuable for trust researchers as more objective factors for potential ride-sharing behavior study in terms of trust-in-platform. Most significantly, we suggest that the factors based on time/location are supposed to be taken into consideration when build up trust-in-platform predict framework. Trust is a premise of ride-sharing platform developing, hence improving the trust between drivers and passengers is a win-win situation.

This study used machine learning algorithm to find that time and location have a relationship with trust in the context of ride-sharing platform and should be regarded as effective factors in evaluating trust-in-platform, which provides a new idea for future studies. And with the information age coming, data mining is more and more popular. This study increased the breadth of related research. Our model may include more variables into consideration and facilitate relevant studies on trust-in-platform by other researchers.

8. Implications, limitations and future research

This study explores whether time and location have relationships with trust in the context of ride-sharing platform, which has several important implications.

Theoretically, a key contribution arises from our focus on the relationship between time/location and trust-in-platform. There have been substantial studies that addressed several antecedents of trust-in-platform

[6][7][8], our research contributes to the existing knowledge of trust-in-platform from time and location perspectives. Moreover, we investigated trust-in-platform influencing factors in ride-sharing platforms, and extend trust-in-platform research in sharing economy business settings, which also contributes to the current trust-in-platform studies in e-commerce transactions [15], in organizational studies [19], and in online social networks [9]. The final contribution of the study is in employing a machine learning approach in ride-sharing studies. Existing studies in sharing economy mostly were conducted through survey [4], interview [9] and experiment [10], advanced machine learning approach corroborates the understanding of our results.

This study also has practical implications. Although time and location are objective factors of travel and are hardly affected by personal will, platform operators could put forwards several solutions to reduce association, thus improve trust-in-platform perceptions to some extends. For example, when the system predicts that the order has a large possibility of matching failure in certain time and location, platform operators could assign drivers with higher skill and ratings to improve the matching rate of orders. Moreover, trust-in-platform prediction model could be used in real business settings. If the trust-in-platform levels are successfully predicted, platform operator could establish reasonable penalties to avoid greater losses. For example, the ride-sharing platform could set low trust-in-platform order with stricter penalty to increase default costs. What's more, if the trust-in-platform levels are successfully predicted, platform operators could establish reasonable penalties to avoid greater losses. For example, the ride-sharing platform could set low trust-in-platform order with stricter penalty to increase default costs. In general, building up trust-in-platform predict framework is a valuable approach to reduce risks and uncertain in the field of sharing economy driven businesses. This research could give clues to the system developer to optimize the rider-sharing platform.

There are also some limitations for this study. In terms of data quality, we hope to get more adequate data to make the model results better. Furthermore, due to the restrictions, it is impossible to obtain the personal behavior and characteristics of drivers and passengers. Hence, it is reasonable that we get not good enough result only depending on the factors of time/location.

Future studies should try to get more effective experiment data with historical behavior of passengers and drivers and more factors should be taken into consideration to optimize the model. In addition, other algorithms of machine learning, including ensemble

learning (Random Forest) and deep learning will be applied. In sum, we hope to build a more complete trust-in-platform prediction framework to improve travel situation in the context of ride-sharing platform.

The future results are expected to include: (1) To identify more factors that influence trust-in-platform related to time and location; (2) To get a more accurate and time-changing trust-in-platform map; (3) To apply deep learning algorithms to predict models and adjust parameters to improve prediction accuracy; (4) To generate a complete prediction model and apply it in real life settings.

9. Acknowledgement

The authors would like to thank National Natural Science Foundation of China (Grant No.71871061) for providing funding for part of this research.

References

- [1] Botsman, R., and R. Rogers. *What's Mine is Yours: The Rise of Collaborative Consumption*, New York: Harper-Collins, 2010.
- [2] S. Benjaafar, G. Kong, X. Li, and C. Courcoubetis. "Peer-to-peer product sharing: Implications for ownership, usage, and social welfare in the sharing economy," *Management Science*, 2018, 65(2), pp. 477-493.
- [3] D.H. McKnight, and N.L. Chervany. "What Trust Means in E-Commerce Customer Relationships: An Interdisciplinary Conceptual Typology," *International Journal of Electronic Commerce*, 2001, 6(2), pp. 35-59.
- [4] F. Hawlitschek, T. Teubner, and C. Weinhardt. "Trust in the Sharing Economy," *Journal of Business Research and Practice*, 2016, 70(1), pp. 26-44.
- [5] E. Ert, A. Fleischer, and N. Magen. "Trust and Reputation in the Sharing Economy: The Role of Personal Photos in Airbnb," *Tourism Management*, 2016, 55, pp. 62-73.
- [6] D. Gefen, E. Karahanna, and D.W. Straub. "Trust and TAM in online shopping: an integrated model," *MIS Quarterly*, 2003, 27(1), pp. 51-90.
- [7] A. Ghose, P.G. Ipeiritos, and B. Li. "Designing Ranking Systems for Hotels on Travel Search Engines by Mining User-Generated and Crowdsourced Content," *Marketing Science*, 2012, 31(3), pp. 493-520.
- [8] X. Cheng, S. Fu, J. Sun, A. Bilgihan, and F. Okumus, "An Investigation on Online Reviews in Sharing Economy Driven Hospitality Platforms: A Viewpoint of Trust," *Tourism Management*, 2019, 71, pp. 366-377.
- [9] S. Chai, and M. Kim. "What Makes Bloggers Share Knowledge? An Investigation on the Role of Trust," *International Journal of Information Management*, 2010, 30(5), pp. 408-415.
- [10] X. Cheng, S. Fu, and G. Yin. "Does Subsidy Work? An Investigation of Post-adoption Switching on Car-hailing Apps," *Journal of Electronic Commerce Research*, 2017, 18(4), pp. 317-329.

- [11] R. Aringhieri, E. Damiani, S.D.C. Di Vimercati, S. Paraboschi, and P. Samarati. "Fuzzy techniques for trust and reputation management in anonymous peer - to - peer systems," *Journal of the American Society for Information Science and Technology*, 2006, 57(4), pp. 528-537.
- [12] R. Zhou, K. Hwang, and M. Cai. "Gossiptrust for Fast Reputation Aggregation in Peer-To-Peer Networks," *IEEE Transactions on Knowledge and Data Engineering*, 2008, 20(9), pp. 1282-1295.
- [13] G. Zervas, D. Proserpio, and B. John. "The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry," *Journal of Marketing Research*, 2017, 54(5), pp. 687-705.
- [14] P. Resnick and R. Zeckhauser. "Trust Among Strangers in Internet Transactions: Empirical Analysis of Ebay's Reputation System," *The Economics of the Internet and E-commerce*, 2002, 11(2), pp. 127-157.
- [15] D.J. Kim, D.L. Ferrin, and H.R. Rao. "A Trust-Based Consumer Decision-Making Model in Electronic Commerce," *Decision Support Systems*, 2008, 44(2), pp. 544-564.
- [16] H. Alon and E. Liat. "The Development of Initial Trust in an Online Company by New Customers," *Information & Management*, 2004, 41(3), pp.377-397.
- [17] X. Cheng, S. Fu, and G.J. de Vreede. "A Mixed Method Investigation of Sharing Economy Driven Car-Hailing Services: Online and Offline Perspectives," *International Journal of Information Management*, 2018, 41, pp. 57-64.
- [18] F. Mazzella, A. Sundararajan, V. D'Espous, and M. Möhlmann. "How Digital Trust Powers the Sharing Economy," *IESE Insight*, 2016, 30, pp. 24-31.
- [19] D.H. McKnight, L.L. Cummings, and L. Norman. "Initial Trust Formation in New Organizational Relationships," *Academy of Management Review*, 1998, 23(3), pp. 473-490.
- [20] X. Liu, G. Tredan, and A. Datta. "A Generic Trust Framework for Large - scale Open Systems Using Machine Learning," *Computational Intelligence*, 2014, 30(4), pp. 700-721.
- [21] Y. Freund and R.E. Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *European Conference on Computational Learning Theory*, 1995, 55(1), pp.119-139.
- [22] Y. Fang, I. Qureshi, H. Sun, P. McCole, E. Ramsey, and K.H. Lim. "Trust, satisfaction, and online repurchase intention: the moderating role of perceived effectiveness of ecommerce institutional mechanisms," *MIS Quarterly*, 2014, 38 (2), pp. 407-427.
- [23] J.D. Cook and T.D. Wall. "New work attitude measures of trust, organizational commitment and personal need nonfulfillment," *Journal of Occupational Psychology*, 1980, 53, pp. 39-52.