Incidence of Online Public Opinion on Guangzhou Simultaneous Renting and Purchasing Policy - A data mining application

Yancheng Wang¹
Haixian Li

Guangdong University of Foreign Studies, China PRC

Abstract

This paper adopts the big data research method, and draws 491 data from the Tianya Forum about the Simultaneous Renting and Purchasing policy of Guangzhou. The qualitative analysis software Nvivo11 is used to cluster the main questions about the Simultaneous Renting and Purchasing policy in the forum. The 36 high-frequency word frequencies are obtained through text clustering. Through rooted theory analysis, the main driving factors for summarizing people’s doubts are 9 main categories, 3 core categories, and the model of driving factors for online forums is established. The study finds that resource factors are the most key factor, economic factors are the important drivers, and policy guiding factors are sub-important drivers.

Keywords: Simultaneous Renting and Purchasing Policy, big data, text clustering, online forum public opinion

¹ All correspondence concerning this article should be addressed to Yancheng Wang at Guangdong University of Foreign Studies, Guangzhou University City, Guangdong, China or by e-mail at wyancheng@yahoo.com.
Introduction

Theoretical basis

Online Forum Public Opinion.

An online forum is an online communication place based on network technology applications. Online forums can generally be divided into comprehensive forums and special topical forums whose functions are teaching, promotion, communication, cultural communication, and so on. Tianya Forum was founded on March 1, 1999. Since its establishment in 1999, the number of registered users has exceeded 100 million, and the daily active user volume has exceeded 1 million, which has great influence on daily life of the Chinese people. The main users of Tianya Forum are Chinese people around the world, who use the network to meet the various needs for communication, expression, and creation among individuals. Tianya Forum has sections for Economics, Real Estate, International Observations, Legal, News Commentary, and others related to the current political situation and the living conditions of the people. People can freely express their opinions in various sectors. Taking the section for Real Estate of Tianya Forum as an example, by the end of August 21, 2018, the number of topics in this section is 720,000, and the number of replies is 9.67 million.

With the increasing use of Internet applications, the establishment of the Internet Forum has opened up a space for the majority of Internet users to express their opinions freely. The Internet has become the main platform for public participation and communication, which has led to the creation of a new online venue, the Online Forum. As the popularity of online forums is increasing, it has now become an important place for online public opinion and one of the most effective communication channels for online public opinion. The online forum embodies the potential value and influence of popular culture and plays an increasingly important role in communication activities. More and more remarks spread through the use of online forums have created a new kind of opinion, the online Forum Public Opinion.

Literature Review

With the rise of major forum websites, online forum public opinion is rapidly developing. Scholars have also conducted relevant empirical research on the use of online forums by the public as political participation. Chen (2015) said that more attention needs to be paid to the growing Chinese media landscape, the various forms, venues of citizenship, political
participation, and the position of personal network in a larger social structure. Chen and Reese (2015) showed from the development of China’s Internet, technology changes and user habits change so that people tend to use different Internet or mobile platforms for different types of communication and interaction. The participation of people through digital communication and media technology will be promoting social changes at many levels and in many forms.

Although less obvious, it has changed in government attitudes and regulatory approaches. Medaglia & Yang (2016) used a theoretical perspective of public scrutiny to analyze the content and responses to more than 18,000 online messages on four controversial topics on Tianya Forum based on evidence from a large number of manually processed vertical data sets. The passage of time, the homogeneity of online discussions, the role of size differences between opinions, emotional polarity, and subject type help to establish a unique system of public review research systems. According to research by Qian (2014), Tianya Forum is one of the most influential Chinese online forums, covering all aspects of social life. There are several characteristics of public opinion communication, which are the popularization of public opinion subjects and the high degree of aggregation of public opinion objects. The public opinion of Tianya Forum often has guiding characteristics.

Yanling, Weihua, and Shuqian (2013) analyzed the "Strong Power Forum" and "Tianya Miscellaneous Talks" to explain the functional characteristics of the political participation of the online forum and the different ways of constructing the public domain, indicating that cyber politics are a new form of democratic political participation, showing a strong social function. However, Xiaoe (2012) finds that there is a lack of rationality in the Internet’s public opinion ecology. The main reason is that the information dissemination makes the information fragmented, which leads to the cognitive dissonance of netizens. The catharsis principle of social psychology negatively produces comments on information. The discussion of the article is based on the online forum about the Guangzhou Simultaneous Renting and Purchasing Policy will also be based on the above research, combined with qualitative research methods and grounded theory to further analyze the driving factors.

Research Methods

Online Big Data Capture

Data mining is the extraction of previously unknown, but potentially useful, and ultimately understandable information and knowledge from a large number of incomplete,
noisy, fuzzy, random raw data. (Jain, 2002). Clustering large amounts of data is a daunting task because the goal is to find the right partition in an unsupervised way, trying to maximize the similarity of objects belonging to the same cluster, and minimize the similarities between objects in different clusters. The online forum has become a source of folk information that governments have given more and more attention in recent years. To get relevant public opinion information from a web forum, you must be familiar with the application of web crawlers. So far, most scholars who have explored big data public opinion analysis mostly use Python language to write programs to extract information from online forums. Figure 1 is a screenshot of a web crawler written in Python by a web crawler we wrote in Python.

**Figure 1** Web crawler code written using Python

### Qualitative Research Method Based on Grounded Theory

Qualitative research method refers to the researcher's delving into the natural situation, through field observation, interviews, recordings, pictures, videos, notes, and other means to reach a comprehensive and in-depth exploration of the nature of things. A theoretical model is generated to obtain a structured explanation of an activity (Xiangming, 2000). Grounded theory is a qualitative research method aimed at developing the basis for the systematic collection and analysis of data. It is not just a coding technique, but also provides a comprehensive theory generation method (Urquhart, Lehmann, & Myers, 2010). This qualitative research method is useful in data collection, analysis, and understanding of results.
Grounded theory emphasizes the construction of theory must be generated from phenomena to essence. We can find out the solution to a problem based on empirical facts, through an in-depth analysis of the data. Since the research on a particular forum based on a particular policy has not yet formed a very complete system, this paper has used qualitative research methods based on grounded theory. First of all, we collate and screen the texts of the Tianya Forum users’ discussion on the Guangzhou Simultaneous Renting and Purchasing Policy. Then, we explore and summarize the discussion texts. Finally, the online forum public opinion drive-driven model is summarized.

**Text Clustering Analysis Method**

Text Clustering Analysis Method is used to divide the data into different groups, and make the gap between the groups as large as possible, and the gap between the data in the group as small as possible. Zhou (2007) believes that when text clustering is performed, the accuracy of clustering results is often low because the text contains too many "general words" and too few "core words" with different degrees of discrimination. A text clustering method based on semantic filtering model is described by Min, Jiajia, Jiahui, Jimin, & Jiping (2015). Text clustering can also be based on frequent item sets in a text set. The basic idea is to treat each text in the text set as a transaction, and to treat the words in the text as data items contained in the transaction. The frequent item set mining algorithm finds a set of items that frequently appear in the text set, and then constructs clusters based on these frequent item sets. This method not only greatly reduces the dimension of the text in clustering, but also provides a clear explanation for the clustering results through frequent item sets.

**Case Study and Data Analysis**

**Case Study**

China started to implement reforms on the housing system in 1998, but housing prices have continued to rise. To address this problem, the government has introduced a number of policies to control housing prices and the intensity and density of these policies has also increased significantly, but housing prices still remain high. With regard to the real estate market which is closely related to the economy and people's lives, as Xi (2017) explained in a government report of October 2017 suggests that we should adhere to the position of "using houses for living instead of speculating" and speed up the establishment of the housing system of multiple-body supply, multiple-channel protection and simultaneous renting and purchasing. The purpose is to maintain residential properties and accelerate the development of the housing rental market in the cities with large net
inflow of population. By creating the housing system of "Simultaneous Renting and Purchasing" as the main direction, the market becomes the main means to meet the multiple-level demand and the government serves as the main platform to provide basic protection. These policies guarantee the basic housing needs of eligible low-income families with housing problems through providing public rental housing or rental subsidies for them.

In 2017, the Department of Building and Housing selected Guangzhou, Foshan, Shenzhen and another 12 cities as the first pilot units of the housing rental policy. As early as July 17, 2017, the Guangzhou government published the "Notice on Accelerating the Development of the Work Plan for the Housing Leasing Market," and proposed 16 specific measures such as safeguarding the rights and interests of both parties, increasing housing supply, and expanding the modern leasing industry. This includes the establishment of state-owned housing leasing companies. However, as Aqiang (2017) said, the Simultaneous Renting and Purchasing Policy has set a higher threshold, and the tenants who qualify can enjoy the public service rights of housing that are more strict. In the process of continuous introduction and the continuous improvement of regulations of the Simultaneous Renting and Purchasing Policy, the public also had many discussions on the online forum and formed a certain scope of public opinion. Tianya Forum is typical of these forums, as discussed below.

With the popularity of the Internet, people have expressed opinions via a convenient and frequently available channel. Most people choose to express their comments on new government policies and specific social phenomena on social platforms such as forums and Weibo. Behind the exponentially rising amount of information, what are the main concerns of the people and what factors are driving them?

Figure 2 shows the main content of this study and the research methods used in each part of the content in the form of a flow chart. The first step is to use the Python crawler software to retrieve relevant discussions and to screen the original materials from the Tianya Forum of the Guangzhou Simultaneous Renting and Purchasing Policy. In the second step, the qualitative analysis software Nvivo11 was used to obtain the main distribution of the users' opinions about the Simultaneous Renting and Purchasing Policy through text clustering. The third step is to establish a theoretical model, which is generated by analyzing the main factors driving the people's opinions through grounded theory. Finally, a drive-driven model for the online forum opinion is established. The fourth step provides relevant reference measures for the Guangzhou Municipal Government as
well as the Tianya Forum Managers to establish relevant warning mechanisms based on the research conclusions.

Figure 2 The framework of the research method

Sources of data.

This study selected Tianya Forum, a mainstream forum in China. The people who use and reasons why people use Tianya Forum to express their opinions are explained as follows. To begin with, the users are centralized opinion leaders and numerous marginalized grassroots people. Tianya users are universally representative of geographical distribution, age composition, cultural level, income level, and industry job distribution. In addition, the forum focuses on deep interaction. Compared with Weibo (equivalent to China’s Twitter) platform based on opinion leaders and WeChat (equivalent to Chinese version of Facebook) decentralized life social platform software, the topic content is more valuable than the speaker’s popularity in Tianya Forum. Furthermore, the public is free to express their opinions in the forum. The participation is high and there is enough data to conduct text clustering analysis.

Analysis sample description.

We captured a total of 28,667 pieces of information in Tianya Forum using the key words “simultaneous renting and purchasing” and “equal rights for renting and purchasing,” using Python written web crawler on July 10, 2018. Most of the information
that does not meet the requirements was screened out according to the policy release time and related content investigation. In this case, 491 valid texts were obtained, of which the number of authors was 57, covering the data of posters and commenters about “simultaneous renting and purchasing” and “equal rights for renting and purchasing.” The average number of comments per post is 12, the maximum number of interactive posts is 97, the average number of author responses is 11, and the number of interactive persons are 62.

Data Analysis

Word segmentation.

We converted the topics in Tianya Forum into structured text forms through crawler programs and text preprocessing. First, the Nvivo11 clustering algorithm was used to cluster and identify the topic, and to find and analyze the nodes and relationship in the texts. Second, we conducted a word frequency query, which lists the words with the highest frequency in the source, and visualizes the results in the word cloud, tree structure diagram or cluster analysis diagram, as shown in Figure 3. Finally, we selected the clustering result with the lowest recognition cost, and used the evaluation function to select the top 20 hot discussions.

Figure 3 Top 20 hot discussions in Tianya Forum
Emotion analysis.

Yoon, Kim, Chang, and Song (2016) proposed a method for analyzing public opinion on online political issues by automatically detecting the polarity of Twitter data. The first stage of the method uses the Lasso and Ridge models of contraction regression to detect the polarity in the tweet. The second stage identifies the main topic through a potential Dirichlet Analysis (LDA) topic model and uses the term sentiment score to estimate the degree of polarity of the LDA topic. We did not adopt the modeling method. Text analysis is a complex process so that human perception of opinions is generally more accurate.

Based on the theory of sentimental dimension, this paper uses the qualitative analysis software Nvivo11 combined with the correction of two coders to analyze the attitudes of posters and followers on the topic of Guangzhou Simultaneous Renting and Purchasing Policy. Nvivo11 has a scoring system in automatic coding query. Each word containing opinions has a predefined score. The content is encoded as a set of viewpoint nodes, ranging from very positive to very negative. Running a coded query can easily identify the content in a hybrid perceptual encoding. After the software analysis, the sentiment analysis of the text is manually checked by two coders to ensure the consistency and accuracy of the coded sentiment analysis.

Based on the emotional metrics and dimension elements of the online public opinion subject, the sentiment analysis method is used to analyze the sentiment trend and dimension elements of the public opinion forum, and to grasp the emotional dimension and emotional state of the words before and after the policy release. The emotional situation of the person is judged by the early warning decision and strategic guidance. For economic factors, the proportion of positive and negative factors is close, and the proportion of neutral factors accounts for the majority. For policy guidance and resource factors, the positives far outweigh the negatives in this case. It can be seen as Figure 4, in the Tianya Forum, the opinion of the policy is optimistic.
As data is encoded, the relationship structure between the subjective category and the core category is basically determined. Using the Metries and Query functions in Nvivo, we can intuitively understand the relationship between concepts through the matrix, which use the already encoded data as the query target, and create the node matrix according to the search conditions. As shown in Figure 5, the 39 most frequently appearing terms are: loans, real estate, house price, house purchase demand, national economy, response rate, finance, property regulation, affordability, income, securitization, housing area, large and medium cities, multi-channels, service platform, supply subject diversification, construction land, developer, liquidity, commercial house, experimental city, related department, residential renovation, leasing diversification, admission qualification, provident fund, educational resource, live, housing insurance, right of use, ownership, land supply, future development, meet the criteria, housing security, rental market. Is this list really needed?
According to cluster analysis, the online forum public opinion drive model can be refined into 9 main categories, namely, educational resource issues, tenant conditions, rental market supply, property market regulation and housing price, taxation and revenue, real estate finance challenges, the housing leasing transaction and supervision service platform, the housing leasing experimental city, and the government’s housing supply system. These are focused on 3 main areas, which are resource factors, economic factors, and policy guiding factors. These factors can be found in Figure 6.

Figure 5 Keyword coding relationship visualization
This paper collects the people's opinion about Guangzhou Simultaneous Renting and Purchasing Policy through online data mining. After carrying out word frequency analysis, text clustering, material coding and theoretical model construction, the research findings are drawn as below. It has basically clarified the driving factors in the logical relationship of the online forum users’ public opinion. The data coding of each driving factor is shown in Figure 7, and is specifically explained as follows.

**Research Findings**

*Figure 6* The online forum public opinion drive model

*Figure 7* Data encoding of online forum users’ public opinion
Material factors.

The resource factor, that is, the scarcity of resources, is one of the important driving factors for online forum users to express their opinions about the Guangzhou’s Simultaneous Renting and Purchasing Policy in Tianya forum, including educational resource issues, tenant conditions, and supply in the rental market. From the data coding, it can be found that the coding reference point of the resource factor accounts for 53.20% of the total codes, and the resource scarcity is the most important driving factor for the online forum public opinion about the policy.

Of the issue of educational resources, according to the coding materials, online forum users are most concerned about the degree, the future development of children, the allocation of educational resources, and so on. In Guangzhou, a first-tier city, the population mobility is strong, so that it attracts a bulk of migrant workers and talents. However, the education of children is one of the most worrying issues for these people. Although “equal rights for renting and purchasing” can rationally allocate educational resources to a certain extent, users of online forums worry that the growth rate of Guangzhou’s excellent educational resources is far behind the speed of people’s demand.

In terms of the tenant in the policy of renting and purchasing, the users of the online forum also indicated that the conditions of the lessee are difficult, and many of the primary schools in Guangzhou are not only strict in terms of admission, but also have many restrictions. Even if they meet conditions for renting a house, it may not necessarily help their child get a degree, in their mind. Therefore, online forum users believe that the short-term purchase and purchase policy can adjust the household purchase qualification and quantity to a certain extent, but they still worry about whether this policy can really promote the effective growth of educational resources.

In the supply of the rental market, most of the users of the online forum said that although the policy has been put forward, the market may not be able to catch up with the implementation of the national policy, and it is likely that the previous “one house is difficult to buy” will become “one room is difficult to rent.” Therefore, the pressure of people to buy a house is indirectly transferred to renting a house; there is no way to solve the problem completely. In addition, for renters, renting houses lacks a sense of presence and security, fearing that they cannot really enjoy the resources of the house or have the right to use it how they want, such as by decorating.
Economic factors.

The economic factor is the most critical driving factor for online forum users to express their opinions about policy in Tianya Forum, including the challenges of property market regulation and housing price changes, taxation and revenue, and real estate finance. From the data coding, it is found that the coding reference point of economic factors accounts for 29.25% of the total codes. The economic factor is the important factor that drives users to express their opinions in Tianya Forum.

On the issue of property market regulation and housing price changes, users of the online forum said that the country began to implement various housing price control measures very early, but certain measures have been very difficult to implement, and so housing prices remain high. Therefore, online forum users are not confident in the policy. They are worried that the housing prices in the school district will be directly transferred to the high rent. The real profit is not for the renter but the landlord. At the same time, people are worried that if the equal rights for renting and purchasing are truly implemented, people who borrowed to buy a house in the era of housing price bubbles will trigger a new round of agitation due to psychological imbalance, and may also lead to economic turmoil.

On the issue of taxation and revenue, property tax reform has been put on the agenda many times by the government. Internet forum users said that taxes and income are the most important factors in regulating the real estate economy. Some of them are worried that the real estate tax reform may impose an additional burden on them. At the same time, some people hope that the Simultaneous Renting and Purchasing Policy and the reform of the property tax system will be implemented as soon as possible, thus cooling the overheated Guangzhou housing market. For the income problem, most people argue that the government can grasp the taxation and income security simultaneously, so as to effectively adjust the balance of supply and demand in the rental market.

Most Internet users are worried about the challenge of real estate finance. Tianya Forum, as a very open online platform, it seems that most of the speculators may not express their behavior and attitude intuitively, which has led to the online forum users basically supporting the country against the speculators, hoping that the housing prices of Guangzhou can return to an ideal state.
Policy guiding factors.

The policy guiding factor is the sub-important driving factor for the online forum users to express their opinions about policy in Tianya Forum, including the housing lease transaction and supervision service platform, the housing leasing experimental city, and the government housing supply system. From the data coding, it is found that the coding reference point of policy factors accounts for 17.55% of the total codes. The policy-guiding factor is an important driving factor for many online forum users to express their opinions about the policy in Tianya Forum.

In terms of housing rental transactions and regulatory service platforms, users of the online forums have expressed very consistent opinions. People believe that although the country is vigorously promoting the development of the rental market, there are many problems in the rental loan business, including insufficient housing, few choices, incomplete public services, and so on. Relatively speaking, the Pearl River Delta region, especially Guangzhou, has a developed leasing market, but many insiders believe that the original personal housing leasing market information is not public, opaque, and information asymmetrical, which limits the passage of the original part of the housing to the rental platform. The housing sources are the key issue to enhance the attractiveness of trading platforms. In addition, most people are also worried that the supervision of the housing leasing market is not strong, allowing speculators to take advantage of the opportunity, but not those who really need it to make a profit.

In terms of housing rental experimental cities, many online forum users have mentioned other domestic housing leasing experimental cities, indicating that other experimental cities also have the policy of equal rights for purchasing and renting. After the implementation of the policy, housing prices have not declined and houses in the neighborhoods near schools are still very expensive. Therefore, they have deep doubts about the effects of policies issued by first-tier cities with unequal supply and demand of educational resources like Guangzhou.

In terms of the government’s housing supply system, most of the online forum users who mentioned this said that the supply of the rental market is insufficient and the application procedures are difficult, which makes it difficult to take advantage of the Simultaneous Renting and Purchasing Policy.
Conclusion

This study shows through qualitative research that the drivers of online forum user policy opinions related to the policy mainly include three core categories which are resource factors, economic factors, and policy guiding factors. In terms of the Guangzhou Simultaneous Renting and Purchasing Policy, the resource factor is the key driving factor, including the three main categories of educational resource issues, tenant conditions, and rental market supply. Economic factors are the most critical factors in the online forum user policy drivers, including the three main areas, which are property market regulation and housing price changes, taxation and revenue, and real estate finance challenges. Policy guiding factors are sub-important in the online forum user policy drivers and three main categories are housing leasing transactions and regulatory service platforms, housing leasing experimental cities, and government housing supply systems.

On this basis, this paper constructs an online forum public opinion driven model. This paper not only goes deep into the coding materials and analysis, but also some expansions are suggested below. This paper particularly emphasizes that the scarcity of excellent educational resources exerts an important impact on the current online forum users’ concerns. With the rapid development of the Internet, the online forum is becoming a new way for the masses to make comments on certain issues, and the government understands people’s opinions and communicates with the public. What is of utmost importance, in this context, is that we figured out that, under the government’s vigorous implementation of various housing price control measures, the policy guidance factors have an important impact on the online forum users’ expression about policy concerns.

Limitation and Outlook

The following research limitations exist in this paper. To start with, at present, only 12 cities in China are piloting Simultaneous Renting and Purchasing Policy. The implementation in each city is not completely consistent. The inconsistency of the relevant rules may lead to inconsistent opinions expressed by users of Tianya Forum, which will result in certain impacts on the implementation of research results. Although Tianya Forum is one of the effective ways for people to express their opinions on the Guangzhou Simultaneous Renting and Purchasing Policy, it does not fully represent the opinions of all residents. Therefore, the extensive sources of the original data of the research are still
lacking. The coding of nodes depends largely on the cluster analysis with certain manual adjustments. It is necessary to rely on the coders’ familiarity with the original materials.

This paper has the following research outlook. First of all, traditional surveys generally require a lot of manpower and material resources, and some of the respondents are unable to express their ideas openly because of psychological influences or other factors. This paper has suggested a good way to avoid the shortcomings of traditional surveys. The data is from an open forum with many users. Through big data collection and data mining, new research methods can be promoted. This paper is a qualitative analysis of the online forum combined with big data methodology. It is innovative to adopt big data as an auxiliary method. Compared with previous research, online big data methodology, as a method of public opinion research, has various advantages. It can not only obtain data in real time, which is quick and accurate, but also help to analyze, monitor and manage online public opinion.

Last but not least; this paper takes Tianya Forum as an example to study the public opinion of Guangzhou Simultaneous Renting and Buying Policy. However, the research methods used in this paper can be extended to scholars and the government to truly improve people's lives. The research methods combine online big data, grounded theory and text clustering, which illuminated the essential from the bottom up phenomenon. It provides a new idea for online public opinion researchers. From the perspective of online public opinion, we can understand the suggestions of the people and adjust the implementation methods of the policies just quickly. It will help the society, the government, and the website administrators to fully understand the public’s opinions about the policy and formulate corresponding policies. We can provide the government with methods to promote the policies and at the same time provide relevant recommendations for the introduction and adjustment of supporting policies. It can be seen that the research method of this paper is not only conducive to making the voice of the people heard by the government, thus helping the government improve people’s lives, but promote the innovation of the research methods of the online forums. Public opinion research conducted online will be more efficient, convenient, and accurate in the future.
References


Aqiang, Z.(2017, July 31). Promote simultaneous renting and purchasing policy and meet the needs of the public, 12 Cities enter the “Anyone who needs to settle down should receive a residence”, China City News, Version 001.


**Biographical Notes**

**Yancheng Wang** earned her bachelor of business administration from the School of Business at Guangdong University of Foreign Studies. Her research interests include social governance, policy studies, data mining, corporate social responsibility.

She can be reached at Guangdong University of Foreign Studies, Guangzhou University City, Guangdong, China or by e-mail at wyancheng@yahoo.com.

**Haixian Li** earned her bachelor of business administration from the School of Business at Guangdong University of Foreign Studies. Her research interests include policy studies, data mining, housing economics, and urban economics.

She can be reached at Guangdong University of Foreign Studies, Guangzhou University City, Guangdong, China or by e-mail at 20151804379@gdufs.edu.cn

Date of Submission: 2018-07-31
Date of the Review Results: 2018-08-13
Date of the Decision: 2018-08-21