Analysis of Text Information of Sharing Economy

Based on the Information Mining of News Information Data

Kai Li
School of Statistics and Mathematics
Zhongnan University of Economics and Law
Wuhan, China

Shujian Xiang*
Zhejiang Gongshang University
Hangzhou, China 310018
*Corresponding Author

Abstract—**The rapid development of sharing economy has drawn the attention of all sectors of society. The academic circle has conducted in-depth study on sharing economy from theory to individual cases. However, there are rarely scholars developing discussions on the understanding of the household sector, the important participant of sharing economy. Accordingly, this paper uses the reports about sharing economy in the internet information to conduct public opinion analysis and following conclusions are obtained through the analysis of topic model of this paper (LDA) and social network: (1) the reports from internet information about sharing economy are basically consistent with the studies of academic circles; however, there are certain cognitive confusions in the public about whether it is sharing economy or sharing conducts; (2) the rapid development of sharing economy poses challenges to the system arrangement; thereinto, the issues of social credit and safety are the main topics that the public focus on; (3) the connotation and development mechanism of sharing economy can be explored through the public opinion analysis based on internet information in an effective mode and it is hopeful to introduce the studies on sharing economy and explore the laws of economics that have not been found out.**

Keywords—sharing economy; natural language processing; topic model; public opinions

I. INTRODUCTION

As a brand new business model, the sharing economy depends on the development of mobile internet technology and gains rapid development and the scale is becoming larger and larger under the trend of ―mass entrepreneurship and innovation" in China. At the same time, all sectors of society become interested in the sharing economy.

In the academic circle, Martin Wiesmann comprehensively proposed the theory of ―sharing economy" in the book Sharing Economy: Using Sharing System to Replace Wage System published in 1984 and thought that the essence of ―sharing economy" is to achieve collaborative sharing of ownership, right of use and right of allocation of capital; Botsman and Rogers (2010) thought that the sharing economy at the present stage is based on the concept of ―Collaborative Consumption which is to link, collect and organize associations through the mode of network connections and conduct effective matching to both the demands of various kinds of sources and the supplier so as to accomplish resource exchange. At the same time, the collaborative consumption possesses four features:

cluster effect, idle capacity, social public resources and the trust among strangers [1]. Botsman and Rogers proposed that the current development state of sharing economy is a kind of typical collaborative consumption economic model. Li Xiaohua (2017) thought that the driving force for sharing economy is from the economic surplus, scarcity of resources, the advancement of information technology and the improvement of ownership. And Li also pointed out that the online ride-hailing service and periodic lease are respectively sharing economy in narrow sense and broad sense. The sharing economy in narrow sense refers to the sharing depending on only spare time or idle resource while the sharing economy in broad sense broadens the boundary [2]. Lu Xianxiang (2016) proposed that the share economy is sharing economy and has features including reduction of transaction costs through platform, optimization of allocation of resources and cognitive surplus [3].

On the official level, the Office for National Statistics thought that the sharing economy is: to share the assets that are not fully utilized only through the mode of realization of ―point-to-point" on the digital platform so as to enable all aspects to benefit from the application beyond of the main purposes of such assets [4]. According to the Development Report of China Sharing Economy 2017 issued by Research Center for Sharing Economy of National Information Center, the sharing economy refers to ―the sum of economic activities using internet and other modern information technology and taking right of use as main feature to integrate the numerous and decentralized resources and meet the diversified demands"[1].

According to the studies of above scholars and official studies, we can think that the sharing economy possesses following features: the first is sharing platform which is to form the supply side of dispersed resources and the demand side of resources into ―multipoint-to-multipoint" associations through the sharing platform based on internet technology; the second is full recycling of idle resources. For the owners, some of the functions of the resources traded in sharing economy have not been utilized and this part of functions is valuable for other consumers. And they may be willing to pay certain remuneration to the owner for use of this part of functions. Therefore, the owner of idle resource will realize the full use of resources when earning profits through sharing such resources;

---

[1] National information center:www.sic.gov.cn/News/568/7737.htm

the third is the transaction mode of temporary transfer of right of use. Different from the ownership of idle assets that is transacted by seller and buyer in traditional economy, what is transacted in sharing economy is the right of use of temporary lease while the ownership of assets is not transferred therewith; the forth is that we can consider the sharing economy as a kind of business model which broadens the boundary of sharing economy.

However, the main participants of sharing economy are concentrated in the household sector. Then the cognition of household sector on sharing economy lacks systematic study as far as this paper is concerned. This paper considers that the cognition degree of households on sharing economy determines their expectations of participation in sharing economy. Under the macro-background where our government vigorously promotes the development of sharing economy, understanding the cognition of residents on sharing economy can help with the decision-making of policy makers and is also favorable for the sharing economy enterprises to formulate the operation plans that can conform to market demands. Accordingly, this paper attempts to conduct information mining to the reports related to sharing economy in internet news through the mining method for text data- based on the big data crawler technique and natural language processing technology so as to analyze the current cognition situation of the public of China on sharing economy.

The following contents of this paper are arranged as follow: part II: research method statement. This party specifies the statistic model and statistical technology applied in this paper; part III: empirical analysis. Explore the ideology of sharing economy in the public through analysis of the relations between hot words, themes and key words in the news information; part IV is the conclusion.

II. RESEARCH METHOD STATEMENT

Text data analysis means to conduct intelligent analysis to the semantic structure formed by specific language vocabulary in text data through computer and excavate the potential relevance in the key works in texts based on Natural Language Processing (NPL) so as to realize the extraction of important information. To analyze the current situation of development of sharing economy in China, this paper uses the fairly popular topic model (Latent Dirichlet Allocation, LDA) in text analysis to extract the main contents of netnews with the title of sharing economy.

A. LDA Model

LDA model is a kind of Generative statistical model [5] proposed by Blei, David M. et al(2003). Its main idea is to explain why some concentrated parts in observation sample are similar based on potential Dirichlet allocation and by using the samples not observed. In the text analysis, it is assumed that the observation samples are the terms extracted from different texts so as to form different topics through combination of above terms and that all the texts are the combination of a few of topics. For a document, we should extract one topic from the topic allocation and extract a term according to the term corresponding to the extracted topic. Repeat the above process

to continuously extract terms until all the terms in this document have been traversed and generated.

Assuming the document $d_i \in D = \{d_i | d_1, d_2, ..., d_n\}$ and D is the document collection, LDA model is set as follow:

- The prior probability of random drawing of a document d_i is $P(d_i)$;

- The topic allocation of document d_i is θ_i and θ_i complies with the Dirichlet allocation with the hyper-parameter of α;

- Generate the topic z_{ij} of No. j word in document d_i through the polynomial of the topic;

- Take samples from the Dirichlet allocation β to generate the corresponding word allocation $\phi_{z_{ij}}$ of topic z_{ij} and the word allocation $\phi_{z_{ij}}$ shall be generated through Dirichlet allocation of hyper-parameter β;

- Generate the final word w_{ij} from the polynomial allocation $\phi_{z_{ij}}$ of the word.

According to the five steps mentioned above, we can divide the generation of a document into two processes: firstly, generate a topic allocation through Dirichlet allocation; secondly, generate the word complying with this topic according to topic allocation. Therefore, the LDA model can be written in the form of simultaneous allocation:

$$P(w_i, z_i, \theta_i, \Phi | \alpha, \beta) = \prod_{j=1}^{N} P(\theta_i | \alpha) P(z_{ij} | \theta_i) P(\Phi | \beta) P(w_{ij} | \phi_{z_{ij}})$$

$$(1)$$

In formula (1): Φ is the unobservable hidden variable.

Integrate θ_i and Φ and conduct summation to z_i and we have:

$$P(w_i | \alpha, \beta) = \int_{\theta_i} \int_{\Phi} \sum_{z_i} P(w_i, z_i, \theta_i, \Phi | \alpha, \beta)$$

(2)

Finally, conduct Gibbs sampling to formula (2) and solve the maximum likelihood estimation.

B. Technical Proposal

Because the text analysis requires technical processing including participle, deletion of stop words, calculation of word frequency and feature extraction, etc, the following research thoughts are designed in this paper:

1) Document collection based on web crawler technology: The purpose of this paper is to understand the cognition of the public on sharing economy and therefore collects the news information with topics including the key word of ―sharing economy by means of crawler technology so as to establish the sample document collection. This paper firstly obtains total 740 news information links through Baidu on the internet and then gets rid of the invalid links and repeated contents and keeps 487 pieces of news information as the sample set.

2) Preprocessing of documents: Because part of the documents collected contain excessive noise (advertisement or the contents different from topic contents), this paper applies text extraction algorithm[2] to extract the core contents.

3) Establishment of word bank including participle and stop words: The important step in text analysis is the extraction of words in the document to prevent the words from being split during extraction process, for example, the ―sharing economy" may be split into ―sharing" and ―economy". Therefore, it requires to predefine the few professional terms. At the same time, the corresponding stop words bank shall be established for the large quantity of meaningless words in the text (such as: preposition, ―of" and ―but", etc) to delete such words so as to achieve the requirement of reduction of noise.

4) Calculating the documents-terms (dt) matrix and establishing LDA mode: Establish the dt matrix with conduct document and rank terms to be used for the topic extraction of LDA model.

5) Optimizing the number of topics to extract the textual features (key words): When the number of topic k is fixed, LDA model will determine the topic term according to the maximum likelihood estimation of term allocation in formula (1). Therefore, the selection of k is the key. This paper refers to the topic selection standard proposed by Arun, et al, (2010) [6], Cao, et al, (2009) [7], Griffiths, et al, (2004) [8] and Deveaud, et al, (2014) [9].

6) Inspecting the relevance of textual features and conduct hierarchical cluster analysis: On the basis of topic terms extracted from (5), this paper conducts relevance inspection to all the featured terms and ―sharing economy" and gets rid of the indistinctive results on the level of 0.05 and conducts

hierarchical cluster analysis to the terms reserved so as to explore the potential connection.

7) Using the social network and the key words involved with the connotation of sharing economy to conduct correlation analysis.

III. EMPIRICAL ANALYSIS

For the sample set constructed by using internet information, this paper firstly analyzes the hot terms (which are the commonly used words); secondly this paper extracts 26 topic features from sample set and conduct correlation test and hierarchical cluster analysis to such features and finally conducts correlation analysis to the key words.

A. Statistics of Hot Terms

Through the calculation of frequency of occurrence of all terms in all the samples, Table I displays the top 20 terms with highest frequency of occurrence.

TABLE I.   TOP 20 TERMS WITH HIGHEST FREQUENCY OF OCCURRENCE

S/N	Term	Frequency of Occurrence (Times)	S/N	Term	Frequency of Occurrence (Times)
1	Sharing economy	8461	11	Field	1350
2	Development	4176	12	Industry	1350
3	Sharing	2413	13	Resource	1314
4	Economy	2071	14	Innovation	1312
5	China	2058	15	Sharing	1253
6	Internet	1918	16	Market	1230
7	Enterprise	1833	17	Society	1136
8	Platform	1828	18	Model	1096
9	Service	1676	19	Provide	971
10	New	1532	20	User	939

In ―Table I", the size of term indicates the degree of corresponding term frequency. Compared with the total 24292 nonredundant terms, the term of sharing appears 2413 times; the term frequency that follows is internet (1918 times), platform (1828 times) and service (1676 times).

Thus it can be seen that such frequently used terms manifest the understanding of the public on sharing economy: sharing economy is the service provided based on internet platform enterprises. Secondly, the terms including consumption, demand, idle, idle resource, business model and innovation, etc manifest that sharing economy is a kind of innovative business model and is manifested as meeting the demands of consumers through transaction of idle resources. However, the terms including sharing economy, sharing and bicycle sharing, etc appearing in the cognition of the public indicate that part of the public may confuse the ―sharing economy" with ―share economy" under the condition of lack of the clear definition of sharing economy.

B. Analysis of Topic of Sharing Economy Document

According to the selection standard for number of topic k in above text, four tendency charts of vibration of standards with k (as shown in ―Fig.1") are drawn in this paper. Under the condition of minimizing Caojuan 2009 and Arun 2010 and maximizing the Griffiths2004 and Deveaud2014 at the same

---

[2]   Refer to https://newspaper.readthedocs.io/en/latest/.

time, the result of ―Fig.2" is displayed between 26 and 30 which is a reasonable value range of topic number. Hereby, this
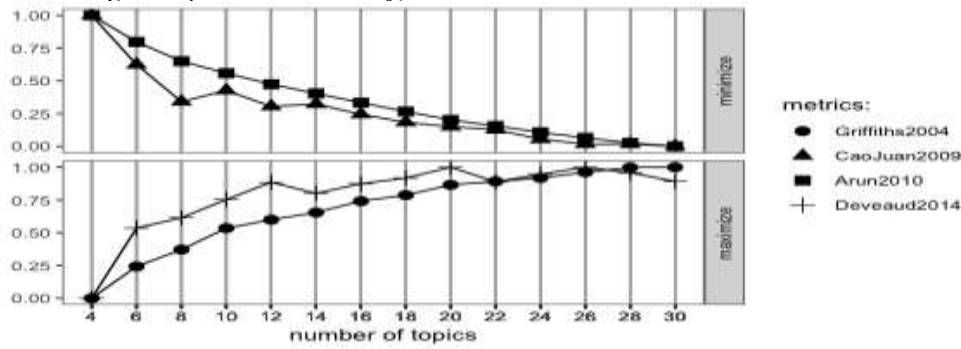
paper selects 26 topics.

Fig. 1. Inspection of standards for selection of number of topic.

Through the analysis of 26 topics, this paper conducts extraction according to the definition of topics and concludes into following aspects which are as shown in ―Table I".

TABLE II. CLASSIFICATION OF TOPIC OF COGNITION ON SHARING ECONOMY

Field	Topic
Management of sharing economy	Credit, system, issue, trust, safety, law and supervision
	Development, innovation, governance, environment, positive, open, promotion, information and mechanism
	Supervision, development, policy, opinion, encouragement, sector, platform enterprise, management and issuance
Connotation of sharing economy	Internet, data, development, industry, new, mobile, integration, technology, manufacturing and transformation
	Sharing, utilization, resource, efficiency, demand, cost, mode, huge and society
	Resource, idle, sharing, idle resource, transaction, cost, article, new and society
Model of sharing economy	Platform, industry, online ride-hailing, management, Didi, appearance, tailored taxi service, private car and ride sharing
	Work, provide, service, relation, house, life, employment, labor and between
	User, service, value, product, experience, provide, demand, Xiaozhu, platform and social contact
	Service, America, company, Uber, provide, taxi, Airbnb, data and economy

It can be seen from ―Table II" that the understanding of the public on sharing economy is mainly concentrated on three aspects: management, connotation and model of sharing economy.

1) Management of sharing economy: The topics focused by the public mainly include credit, safety, supervision and the policy encouragement of government sector on sharing economy. The development of sharing economy poses the requirement of reform of the current social system and the reduction of transaction costs incurs the ever increasing separation of right of use and ownership of transaction goods. The method for guarantee on the interests of both transaction parties shall be ensured on the system level. However, the

time of development of sharing economy in China is relatively short and the formulation of corresponding supervision policies and the issuance of management regulations cannot follow the current development situation of sharing economy which causes the hidden dangers in credit and safety; secondly, the reasonable system arrangement is the external condition for benign development of sharing economy and the complete legal system is the key to guarantee the sharing economy with the cognitive surplus as driving force to continuously innovate by integrating the traditional economy.

2) Connotation of sharing economy: The right of use of idle resources is deemed as the transaction object of sharing economy which promotes the innovation and industrial upgrading of business model. Unlike the information asymmetry in the transaction of traditional market, the enterprises on sharing platform integrated with internet and mobile communication technology concentrate the scattered transaction information in market and form a market with information symmetry. In such market, the information mastered by both transaction parties is complete which reduces the agent costs and realizes the effective allocation of resources and extends the social resource stock boundary. For the traditional enterprises, the information symmetry of sharing economy can relieve the issue of idle capacity and further drive the enterprise innovation and transformation so as to incur the transformation of national economy industry.

3) Model of sharing economy. According to the current development of sharing economy, the public think that the online ride-hailing (ride sharing, Didi and Uber) and the short-term rent (Airbnb and Xiaozhu) provided based on sharing platform are the hot sharing modes. At the same time, in the sharing economy, the residents can use the idle resources or spare time to gain profits, such as private cars, knowledge and skills, etc. In the spare time, the residents can provide services to the demanders of those resources by means of their cognitive surplus and production skills. The above behaviors change the traditional economic production modes and customer behaviour. Therefore, the sharing economy will impact the employment and means of labor of household sector.

C. *Hierarchical Cluster Analysis of Document Key Words*

According to hierarchical cluster analysis, this paper conducts cluster to the key words related to the significance in ―sharing economy". And the results are as shown in ―Fig.2".
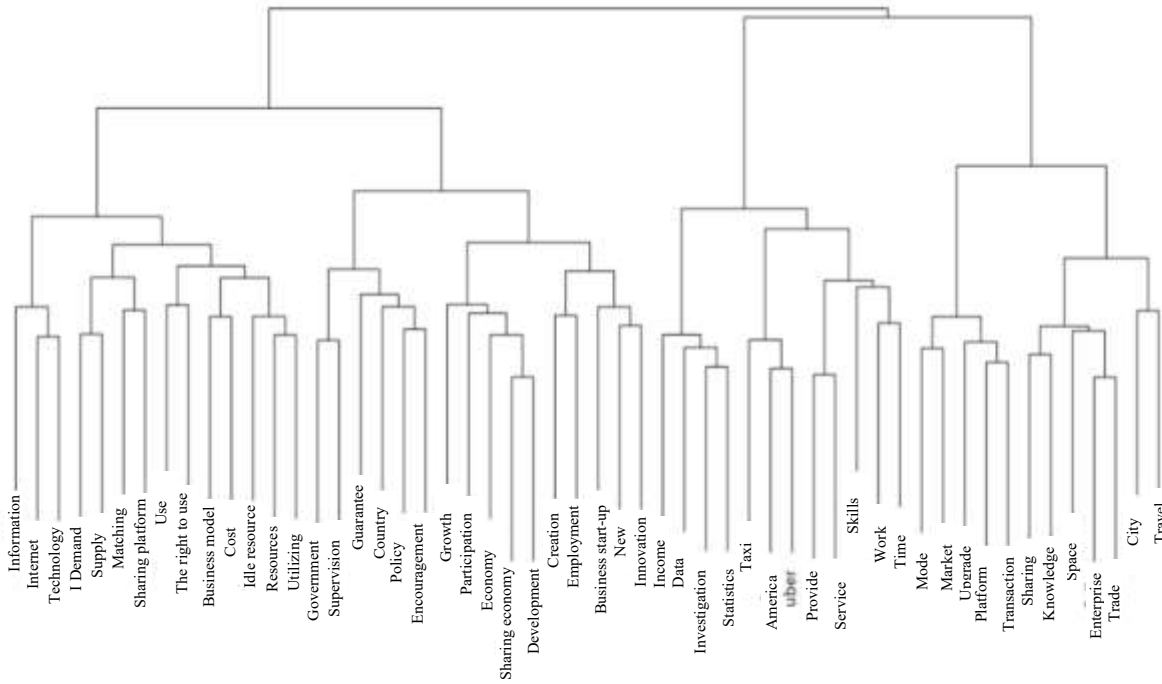

Fig. 2. Hierarchical cluster analysis of key words.

According to ―Fig.2", the hierarchical clustering algorithm classify the semanteme related to internet news information related to sharing economy into 2 main classifications and four detailed classifications.

First main classification includes two detailed classifications including connotation of sharing economy and the macro-development situation. In the detailed classifications, ―information-internet and technology" indicates that the transaction information in sharing economy is attached to the network communication technology. ―Demand and supply — matching and sharing platform" explains the transaction model of sharing economy which means that the transaction is realized through sharing platform by matching the demands and both the supply and demand sides. ―Right of use and use- commercial model and cost- idle resource-resource and utilization" describes that the sharing economy releases the resource values that have not been fully utilized through transaction of right of use of idle resources and the production costs and consumption costs are reduced under this business model.

In the detailed classifications of macro-development situation of sharing economy, ―government and supervision — guarantee, state, policy and encouragement" indicates that the Chinese government encourages the development of sharing economy at the present stage and there are also the requirements of supervision and guarantee. ―Growth — participation — economy and development — creation — creation and employment -entrepreneurship and Innovation"

indicates that the innovation incurred by sharing economy drives the employment and promotes the growth of social economy.

Second main classification mainly includes the practice contents of sharing economy. ―Income — data and statistics, investigation — taxi — America, Uber" indicates that the services provided in sharing economy are from the individual skills and time; at the same time, ―model and market — upgrading — platform and transaction" explains how the sharing economy upgrades the business models in economic market. ―Sharing and knowledge — space — enterprise and industry — city and appearance" further indicates the specific realization path of economic business model of sharing economy where the sharing economy appears in cities and is promoted and applied in national economy industry and is involved with the sharing of knowledge and space.

According to the above analysis, we can think that the text mining of sharing economy can effectively extract the key information and can reflect the cognition of the public on the development of national economy in a relatively correct mode.

IV. CONCLUSION

This paper conducts information mining to the internet information text of sharing economy and indicates following

according to the hierarchical clustering of topics and key words and the results of social network analysis:

- Currently, the cognition of the public on sharing economy keeps consistent with that of academic circle to a great extent. From the results of topic analysis, the sharing economy known to the public is mainly concentrated on: new business models, utilization of idle resources and matching mechanism of sharing platform. Thereinto, the online ride-hailing and short-term rent of house are the two kinds of sharing business models that are the most familiar to the public. In addition, there are ambiguities in the differences between share economy and sharing economy of the public which indicates that whether it is share or sharing is still an ambiguous concept to the public. To this end, it is helpful for the public to distinguish them to define the boundary between share economy and sharing economy.

- Chinese government has shown positive promotion attitude to the sharing economy and the public mainly pay attention to the issues including credit and safety, etc incurred by sharing economy. In recent years, the Chinese government has proposed the strategy of sustainable development of society. However, there is deviation with the current development situation and the operation mechanism of sharing economy to a certain degree. The sharing economy realizes effective allocation through integration of social idle resources. But there are also the conflicts of game on multiple aspects. The solving of the demands and contradictions of various parties cannot be realized without the support of system arrangement. Therefore, the government shall accelerate the establishment of credit mechanism and the formulation of safety regulations in sharing economy to eliminate the shortage of related legal guarantee for sharing economy.

- The text mining technology can effectively extract the reports related to sharing economy so as to obtain the valuable key information. According to the hierarchical cluster analysis on textual features, it is thought in this paper that currently the sharing economy is the economic form which takes the idle resources as core and matches the provider and consumer of right of use of idle resources in the market based on the information transmission technology of internet and through the sharing platform and by means of the advantage of lower transaction costs than that of traditional transaction costs to improve the efficiency of social resource allocation.

REFERENCES

[1] Rachel Botsman and Roo Rogers. What's Mine Is Yours: The Rise of Collaborative Consumption [J]. Harper Collins, 2010.

[2] Li Xiaohua. Connotation and Feature Analysis of Sharing Economy [J]. Business Study, 2017, 59 (7): 119-126. (in Chinese)

[3] Lu Xianxiang. Sharing Economy: Minimization of Transaction Cost, System Reform and System Supplies [J]. Social Science Front, 2016(09):51-61. (in Chinese)

[4] Hardie M. The feasibility of measuring the sharing economy[R]. Office for National Statistics. 2016.

[5] Blei, David M., Ng, Andrew Y. et al. Latent Dirichlet Allocation. Journal of Machine Learning Research[J], (2003) 3 (4–5): pp. 993-1022.

[6] Arun R, Suresh V, Madhavan CV, Murthy MN. On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations[R]. In Pacific-Asia Conference on Knowledge Discovery and Data Mining 2010 Jun 21 (pp. 391-402). Springer, Berlin, Heidelberg.

[7] Cao, Juan, et al. A Density-based Method for Adaptive LDA Model Selection[J]. Neurocomputing 72.7-9 (2009): 1775-1781.

[8] Deveaud Romain, Eric SanJuan and Patrice Bellot. Accurate and Effective Latent Concept Modeling for Adhoc Information Retrieval[R]. Document numérique 17.1 (2014): 61-84.

[9] Griffiths Thomas L. and Mark Steyvers. Finding Scientific Topics[R]. Proceedings of the National academy of Sciences 101.suppl 1 (2004): 5228-5235.