The opinion of crowd participatory for OpenStreetMap: a survey in China

Jie Shen[1, 2, 3,*], Shuai Yang[1, 2], Kaiyue Zang[1, 2], Jiemin Zhou[1, 2], Hongchao Fan[4,*] Bo Mao[5]

[1]Key Laboratory of Virtual Geographic Environment (Nanjing Normal University), Ministry of Education, Nanjing 210023, China
[2]School of Geography Science, Nanjing Normal University, Nanjing 210023, China
[3]Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
[4] School of Remote Sensing and information Engineering, WUHAN University, Wuhan 430079, China - hongchao.fan@whu.edu.cn
[5] Collaborative Innovation Center for Modern Grain Circulation and Safety, Jiangsu Key Laboratory of Modern Logistics, Nanjing University of Finance and Economic, Nanjing 210003, China

ABSTRACT:

At the present, OpenStreetMap (OSM) is considered as one of the most successful and popular VGI (Volunteered Geographic Information) projects. It provides a platform that all the registered members coming from different areas in the world can cooperate with each other to mapping our world. Besides, OSM attracts more and more people, companies and even the governmental agencies because of its free and open source. Studies have proofed that both the quantity and quality of OSM data in several western countries, i.e. Germany, France and the Netherland are even better than the authority data. In recent years, the quantity of the OSM data and the number of contributors in China increased rapidly, but the overall distribution of OSM data is very fit with the distribution of population and the economic development and it displays an uneven development in different provinces and cities in China. Besides, the state of the OSM in China is just similar to that in Germany in 2010 in terms of data quantity and quality, although China is about 25 times to Germany regarding land area and the smartphone penetration in China and Germany does not have a large distance (51.7% to 68.8%). Why is the development of OSM in China so poor and backward when comparing that with western countries, although the environment in hardware and software in China are similar to the western countries? Attempting to answer this question, this paper presents a user survey in China. Mainly, knowledges and experiences about OSM and OSM contribution were asked in the user survey. The user survey was conducted both by paper and pen and by using online platform. Totally, over 1200 participants with the age range from 15 to 80 and a huge diverse of background took part in the user survey. In this paper, we would like to describe the design of the questions for the user survey at first. Then we will demonstrate the results of the user survey, as well as the analysis and conclusions, which can be drawn from the user survey.

1. INTRODUCTION

As a famous project of Volunteered Geographic Information (VGI), OpenStreetMap (OSM) has developed rapidly in recent years. It provides a platform that all the registered members coming from different areas in the world can cooperate with each other to mapping our world. For example, OSM has provided detailed map of Haiti in short time to support the rescue after the earthquake in Haiti in 2010. Up to now, there are more than 4.2 million registered members who make OSM growing rapidly (Stereo, 2018). Because of its free and open source policy, more and more people and companies choose OSM. Apple Inc. and Foursquare have replaced Google map with OSM map providing the map service for them after 2012.

The major mechanism behind of this successful story is the development of crowd participatory in the context of Web 2.0 and the advanced technologies in positioning embedded in mobile devices. Studies have proofed that both the quantity and quality of OSM data in several western countries, i.e. Germany, France and the Netherland are even better than the authority data(Dorn, Törnros, and Zipf, 2015; Jokar Arsanjani and Vaz, 2015; Vaz and Jokar Arsanjani, 2015).

Although OSM has been successfully used in the western developed countries, especially in Europe, it seems that the OSM is still an infant in China(Zheng and Zheng, 2014). It may be many factors resulting in this situation.

In recent years, the quantity of the OSM data and the number of contributors in China increased rapidly, but the overall distribution of OSM data is very fit with the distribution of population and the economic development and it displays an uneven development in different provinces and cities in China. Why is the development of OSM in China so poor and backward when comparing that with western countries, although the environment in hardware and software in China are similar to the western countries?

2. STATE OF ART OF OSM IN CHINA

2.1 The current research of OSM in China

China is a country with vast territory and large population, so it is an ideal platform for OSM data collected. But the quantity, quality, application and research of OSM are still under development in China. First of all, we want to reveal the use of OSM in China from the academic research field. So, we searched the thesis and paper in CNKI (China National Knowledge Infrastructure) and Google scholar published during 2010-2015 by Chinese research institutions, using the keyword "OpenStreetMap", at last we found 42 research papers and thesis.

In order to find out the main research topic of those papers, we first extracted the keywords from those papers and made a list

of them. Then we merged all the keywords and abstracts from those papers in one text file. After that, we matched the keyword in the list one by one with the above text file and calculated the frequency of the keyword appeared. We sorted the keywords by the frequency and filter out the keywords, which were seldom appeared. At last we got 18 keywords.

Trying to find out the topic of Chinese research in OSM, we visualized the keywords in Figure 1. First we referred to the process definition of GIS and divided the 18 keywords into four categories, which are located in the middle circle shown in the Figure 1. The outer circle represents the 18 keywords and the length of the arc, which the keyword is located in, is the ratio of the frequency of that keyword. From the Sector size, we can find that the keywords "Road Networks", "WebGIS" and "Navigation" occupy a big proportion. We can infer that there are more researches on those directions than others, such as the data of OSM acquisition and application.
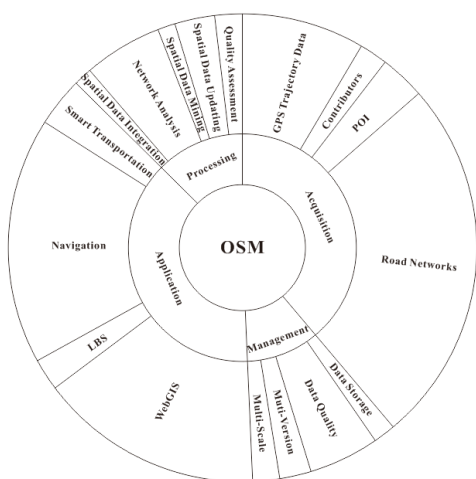

Figure1. The frequency of the keywords in the paper related OSM in China

2.2   The quantity of OSM in China

2.2.1   Comparison OSM data volume of China with other countries

In order to intuitively reflect the gap of the amount of the OSM data between China and other countries, we used a program to extract the number of monthly OSM nodes from January 2007 to December 2015 in China, Germany, U.S.A., Australia and other 8 countries, resulting in Figure 2. The absolute number of OSM nodes has been normalized by the population density in each country to reduce the impact of the size of the country on the result. And in order to reduce the influence of the uneven data gap to the display effect, the vertical axis uses a logarithmic scale. The country names appear right side of Figure 2 in descending order based on the value retrieved from the datasets.
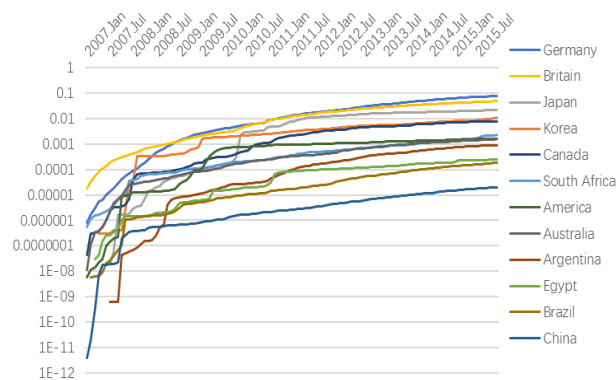

Figure 2. Number of OpenStreetMap (OSM) Nodes per Population/Area-ratio (Jan. 2007–Dec. 2015)

The diagram shows that the amount of OSM data has been greatly enriched in the past nine years. Especially, Germany, Britain and Japan have higher values in comparison to other countries that were tested when considering the relation between the number of OSM members and the total population density. Although the total amount of OSM data in China is not less than that in the other countries, China just row in the end after considering the population density. And this also shows that China has a great development space and potential in the OSM.

2.2.2   The quantity of OSM in different regions in China

In order to display the distribution of OSM nodes data in China, we planned to show it on the administrative map of China. Firstly, all the OSM data of China had been downloading from http://download.geofabrik.de/ in August 14, 2016. The total amount of OSM nodes in China is 169989 (not including Taiwan province). Secondly, Density Analysis Tool in ArcMap 10.2 was used to process the Nodes data and got the density distribution of the nodes data. In order to clearly display difference of nodes data distributed in different areas, we used the method of geometric interval classification, and the data density was divided into nine levels. The selection of color is red to blue band to highlight difference of the data density. Thirdly, the map of Chinese administrative division was overlaid with the data density distribution map after reunification of the projection coordinates of the two maps. Finally, we input the superposed map to ArcScene 10.2 to 3D draw. The elevation is defined as 0.0021 times as much as the value of the data density because the difference of the data density is too large. And then we got Figure 3. The higher the terrain is, the greater the data density is; the flatter the terrain is, the smaller the data density is.
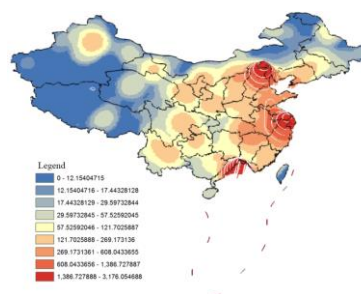

Figure 3. 3D heat map of OSM data density distribution in China

It is obvious in the Figure 3 that the OSM data of China are mainly distributed around the provincial capital cities, especially in Beijing, Shanghai, Guangzhou and Shenzhen. And from the overall perspective, the data in the eastern are significantly larger than that in the central and western. The overall distribution of OSM data is very fit with the distribution of population and the economic development characteristics of various provinces and cities in China.

2.3 The characteristic of the contributors in China

2.3.1 Change in the number of OSM data contributors in China

Although the total number of OSM members in an area might partly reflect the potential data contribution that could occur in an area, it does not take the physical truth of the members into consideration. Therefore, we divided the contributors into three groups according to the number of contributing nodes: who have created fewer than 10 nodes as "Nonrecurring Mappers", who have created more than 10 nodes but fewer than 1000 nodes as "Junior Mappers", and who have created more than 1000 nodes as "Senior Mappers" according to the methods mentioned in (Neis, Zielstra, and Zipf, 2013). We used a program to extract Chinese monthly contributor IDs and their respective number of contributed nodes between January 2007 and December 2015 from the OSM historical data. And then in accordance with the above method, we have made statistics on the number of the three different groups in each half-year, thus we got Figure 4 Number of contributors and distribution of mapper groups.
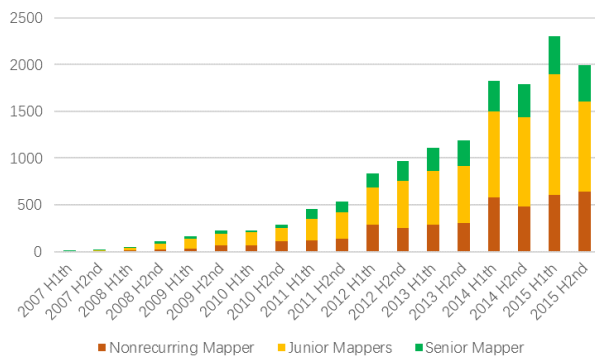
Figure 4. Number of contributors and distribution of mapper groups

The Figure 4 shows that the number of contributors increased year by year and the growth rate is very fast. The number of Junior Mapper is the most in the three groups, and it is in accord with the normal distribution.

2.3.2 Changes in amount of OSM data contribution in China

The data quantity of OSM dataset in the specific area partly reflects its data quality. Thus we extracted the number of Nodes, Ways and Relations from OSM history data of China between 2007 and 2015. And then we got Figure 5 the total number of Nodes, Ways and Relations collected in Chin (2007~2015). In the nine years, whether the growth rate of Nodes, Ways or Relations is very fast. This is consistent with the growth rate of contributors in the front. This partly reflects that the OSM data

accuracy of China is getting higher and higher. At the same time, we can see that the growth rate of the three kinds of data began to flatten. In combination with Figure 3, it is not difficult to know that this is because the amount of data in large cities is gradually saturated, while the contributors in small cities and remote areas are too few.
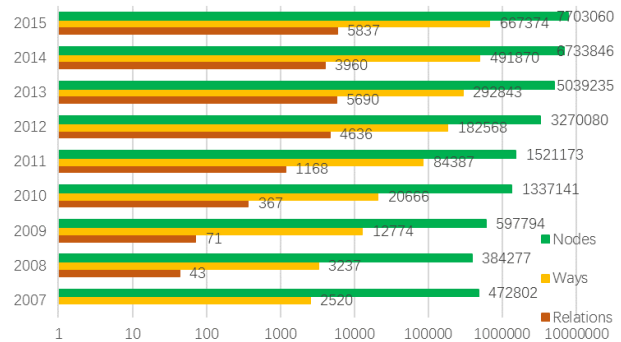
Figure 5. Number of Nodes, Ways and Relations in China (2007~2015)

3. IMPLEMENTATION

There are lots of approaches about the contributors' feature of OSM these years (Stereo, 2018). We can also download OSM spatial data (Geofabrik, 2016) and registered people information (AltogetherLost and Neis, 2018) easily. Neis, Mooney and zielstra analysis of the behaviour of OSM registered users(Mooney and Corcoran, 2014; Neis and Zipf, 2012; Zielstra et al., 2014).

But there are no approaches about unregistered users feature (especially in China). This article will investigate this question. Semi-structured interview and questionnaire method are both used to get people's information (Feng, 2005). We chose questionnaire method in this case considering of saving time and money, anonymity, personal error avoiding and easy counting.

3.1 How and why we designed the questions

The purposes of this research are to investigate the characteristic of people who use OpenStreetMap in life and the behaviour feature when they use them. In order to obtain the user characteristic data, commonly used questionnaires or structured access method. Due questionnaire method can limit the results to options and avoid the investigators' personal tendencies. We decided to adopt the questionnaire.

The user survey was conducted both by paper and online surveying platform. Mainly, personal background, knowledge and experiences about OSM and OSM contribution were asked in the user survey and in total, we set up 22 questions. The workflow of the user survey is as the following. Figure 6 shows the flow chart: (1) All the participants will answer the six questions about personal background, and then will be asked the seventh question which is "Have you heard about OpenStreetMap ever before?" (2) Based on the answer of participants for seventh question, we divide the participants into four categories: *Never heard about OSM, Heard about OSM but never used, Used OSM but never contributed, Used and contributed OSM.* (3) For the first category of participants, we will ask them to try to use OSM at few minutes and then ask

them whether or not willing to use the OSM in the future. (4) For the second category of participants, we will ask them two or three questions about some information about OSM such as the merits and drawbacks of OSM and the open source of OSM. And they also are asked whether or not willing to use the OSM in the future. (5) For the third category of participants, we will ask them how about the interactivity of OSM and whether or not willing to recommend OSM to others besides the questions in (4). (6) For the fourth category of participants, we will also ask them some questions about the edit environment of OSM for user to contribute geodata besides the questions in (5). (7) In the end, all the participants will be asked to give any suggestions about the user survey.
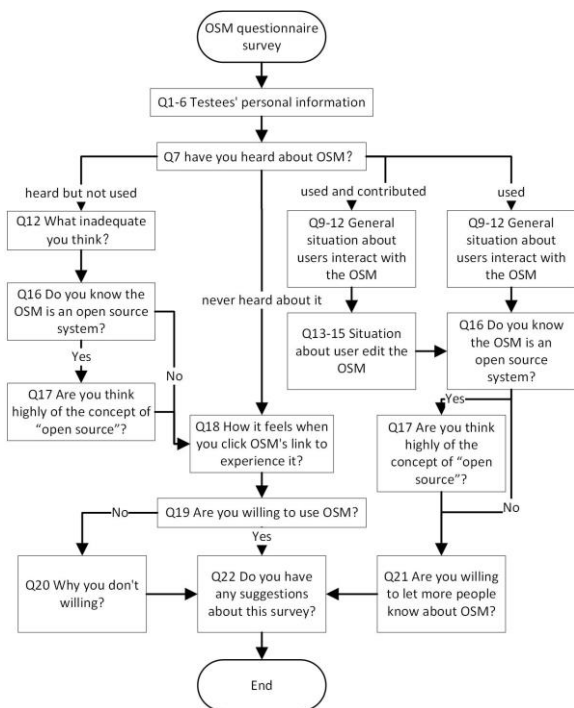

Figure 6. The flow chart of questionnaire that people fill in

3.2 The conduction of the survey

The survey questionnaires were mainly reviewed by four faculty members from four different universities. The universities are Nanjing Normal University, University of Heidelberg, Nanjing University and Nanjing University of Finances and Economics. The questionnaires were pretested with several graduate students who have GIS experience in Nanjing Normal University. Finally, the survey, which was implemented in a Chinese online questionnaire survey website called WJX (www.wjx.cn), was open at 18:55 in April 25, 2016.

The person who was interested in this survey can fill in the questionnaire by scanning QR code or link to the website of the survey. Some fashionable social media tools such as WeChat and QQ were used to spread messages with URL and QR code to the online survey embedded in it. We also made some posters to spread this questionnaire and placed them in canteen, residential areas and supermarkets in the campus. A total of 1,007 responses were received until 18:00 in July 4, 2016. There are 988 valid questionnaires among them.

4. RESULTS AND DISCUSSION

4.1 Questionnaire results

Totally, over 1000 participants with the age range from 15 to 80 and a huge diverse of background took part in our user survey, and we have got 988 effective questionnaires in the end. Most of the respondents are young adult undergraduate students in China. Among 1,007 contributors who took part in the study, most of them are students (64%), 15% are geography or computer science related professionals, 9% are other fields' employees. The remaining 12% respondents are from other occupation. About 90% of the respondents are less than 40 years old, 55% are between 20 and 29 years old, 21% less than 20, 13% are in the range of 30-39, another 8% are between 40-49, and 3% above 50 years. For respondents' sex, the numbers of male and female are similar: 51% are male, and 49% are female.

Because of the high proportion of students, more than half (57%) of the respondents' monthly income less than 2000 CNY. 12% of the respondents' monthly income more than 10000 CNY, 6% are in the range of 8000-9999, 7% are in the range of 6000-7999, 9% are in the range of 4000-5999, last 9% are between 2000 and 3999. About three fifth of the respondents have college/university degrees (62%), 22% have/reading postgraduate degrees, and 12% have/reading Ph.Ds., only 4% have high school or lower education. We believed that some of them, who selected college/university option, are undergraduate student in school on the social context of China. Respondents chose other options also are the same.

Factors		Willingness of using (%)		Willingness of recommending (%)	
		C1	C2	C3	C4
Gender	Male	75	91	86	91
	Female	73	94	87	100
Age	< 20	69	94	86	100
	20~29	77	91	89	100
	30~39	65	94	89	72
	40~49	73	95	55	100
	≥50	100	86	100	100
Educational background	High school or less	77	88	100	-
	Undergraduate	73	95	80	88
	Master	76	94	88	100
	Ph.D.	85	83	91	91
Occupation	Geography or computer science related professional	85	94	93	92
	Company staff	72	89	83	50
	Government staff	80	84	88	-
	Freelancer	64	88	78	-
	Student	73	92	83	100
	Others	74	86	91	100
Monthly income	<2000	72	92	84	100
	2000~3999	79	90	100	100
	4000~5999	78	100	58	100
	6000~7999	78	83	90	100
	8000~9999	83	86	94	100
	≥10000	69	94	89	82

Table 1. Cross effects of some factors on willingness to use and recommend OSM

4.2 Analysis and discussion

Based on the answer of participants for seventh question, we divide the participants into four categories: Never heard about OSM (C1), Heard about OSM but never used (C2), Used OSM but never contributed (C3), Used and contributed OSM (C4). We found that the ratio of the four categories of participants is 22:8:5:1. It can reflect that the number of the users of OSM in China is very low and the number of the contributors is much lower. In order to clearly show the cross-effects between participants'' characteristic variables and the willingness of using OSM of the first two categories (C1 and C2) and the willingness of recommending OSM of the latter two categories (C3 and C4), this paper produced Table 1.

As a whole, the first two types of respondents are highly receptive to OSM. About 79% of participants in the first and second category are glad to use OSM after we introduced it to them. Of course, there are still differences in the willingness to use OSM of the first two types of respondents. Compared with the first category, the second category is more willing to use OSM in the life. This may be because the second category has more exposure to the advantages of OSM when others introduce OSM to them. This is consistent with the willingness to recommend of the last two categories. Meanwhile, the willingness to use OSM of the respondents was also significantly related to the occupational background. Regardless of whether they were the first category or the second category, those who were professional Geography or computer science related professional were most willing to use OSM. The occupational background makes them more willing to try new techniques and products in the field. The second largest group who are willing to use OSM is student. Students can be said to be the group that has the most time and energy to research and use new things.
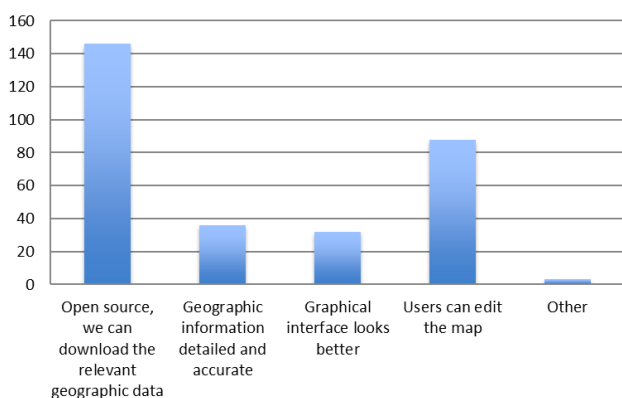

Figure 7. The advantages of OSM compared with other web maps

Similar to the situation of willingness to use OSM, over 87% participants in the third and fourth category are willing to recommend OSM to others and respondents who conducted map editing on OSM are more likely to recommend OSM to others. This matches the statistical result of the 11th question. As shown in Figure 7, most people think that the main advantages of OSM are its open source features and editable. This is the most important reason why OSM is recommended for use. In addition, by analysing the purpose of using OSM, it is also possible to find out the current OSM user groups in China. As shown in Figure 8, the main purpose of using OSM is

to work or learn. In other words, geographically related practitioners or students are the main users of the current OSM in China. Fortunately, some people (9.32%) are more actively involved in the contribution of geographic information. It is believed that as people's awareness of information increases, more and more people will change from users of geographic information to producers of geographic information and contribute to the informationization of life. Meanwhile, due to OSM data coming from different sources, the quality of OSM data in China is uncertainty which makes OSM be in the dry tree among the competitions with other mature platforms of web map such as Baidu and Amap in China, although the open source of OSM data attracts many people to use and spread it. As shown in Figure 9, the main drawback that users report is that the quality and quantity of OSM data cannot be guaranteed. Besides, in the field of data security, contributors may leak out the sensitive information through the OSM, which is an unauthorized platform by the Chinese government.
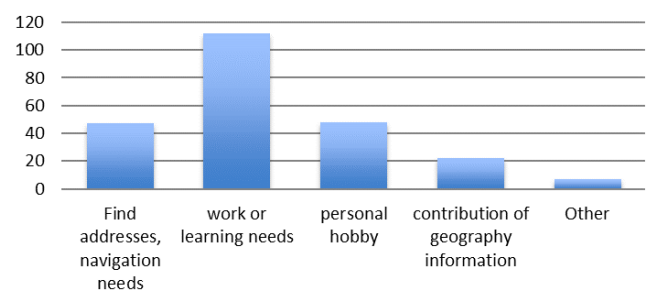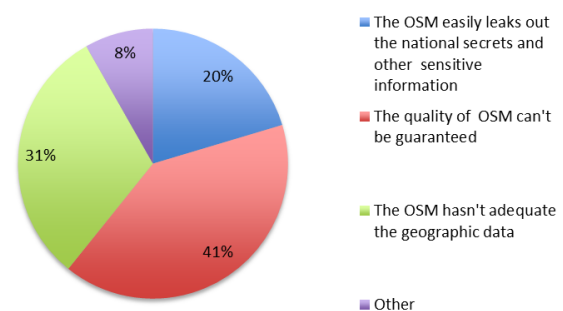

Figure 8. The purpose of using OSM

Figure 9. The inadequate of OSM in users' eyes

5. CONCLUSION

As shown in the above section, we briefly introduced the status of OSM in China, including research areas, data quantity, and characteristics of contributors. During the investigation, it was found that due to the difficulty of direct statistics, there was a lack of understanding of the general users of OSM, especially a large number of unregistered users. Therefore, we designed this user survey and hoped to outline the awareness of OSM in China. The questionnaire design focused on the characteristics of respondents and their understanding of OSM. Because the respondents come from different regions in China and have different educational backgrounds, the questionnaire results can reflect the user group characteristics of OSM in China within a certain aspect.

Based on the questionnaire results, we come to the following conclusions:

1. Only a few people used OSM, and fewer people contributed.

Of the 988 valid questionnaires we received, only 175(17.7%) respondents had used OSM. Further, only 27(2.7%) people contributed data. This shows that OSM is still a very niche map platform in China, which is far less than commercial web maps such as Amap and Baidu.

2. Users are optimistic about OSM open source and are willing to use and promote.

Although most respondents did not know OSM before, through their brief introduction, they also showed great interest in OSM. Recognizing the open source features of the OSM platform is its unique advantage. Therefore, they will use OSM as a supplement, where commercial network maps lack data.

3. In China, Data quality and user experience is not as good as other web maps.

By comparing OSM with other maps, the respondents believe that in economically developed and densely populated areas, the quality of OSM data is not inferior to other maps, but also no obvious advantage. And they are already accustomed to using existing commercial web maps (through the design optimization for the Chinese market, the user experience is significantly better than OSM).

4. No client or App, unfriendly editing interface.

Lack of clients (especially mobile apps) has caused inconvenience to users, raises the barriers to use for ordinary users, and leads users not willing to use the OSM platform. At the same time, the unfriendly editing interface has also deterred the users' willingness to contribute.

In the future, in order to increase the number of users and data quality of OSM. We will further evaluate the quality of POI in OSM platform. And attempted to develop a plug-in for uploading the collected road roughness information as a road attribute information in the OSM database.

ACKNOWLEDGEMENTS

REFERENCES

AltogetherLost and Pascal Neis. 2018. "OSMstats – Statistics of the Free Wiki World Map." http://osmstats.neis-one.org (March 17, 2018).

Dorn, Helen, Tobias Törnros, and Alexander Zipf. 2015. "Quality Evaluation of VGI Using Authoritative Data—A Comparison with Land Use Data in Southern Germany." *ISPRS International Journal of Geo-Information* 4(3),pp. 1657–1671.

Feng, Xiaotian. 2005. *Modern Social Survey Method*. Wuhan: Huazhong University of Science & Technology Press.

Geofabrik. 2016. "OpenStreetMap Data Extracts." http://download.geofabrik.de (March 17, 2018).

Jokar Arsanjani, Jamal, and Eric Vaz. 2015. "An Assessment of a Collaborative Mapping Approach for Exploring Land Use Patterns for Several European Metropolises." *International Journal of Applied Earth Observation and Geoinformation* 35,pp. 329–337.

Mooney, Peter, and Padraig Corcoran. 2014. "Analysis of Interaction and Co-Editing Patterns amongst OpenStreetMap Contributors." *Transactions in GIS* 18(5),pp. 633–659.

Neis, Pascal, Dennis Zielstra, and Alexander Zipf. 2013. "Comparison of Volunteered Geographic Information Data Contributions and Community Development for Selected World Regions." *Future Internet* 5(2),pp. 282–300.

Neis, Pascal, and Alexander Zipf. 2012. "Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap." *ISPRS International Journal of Geo-Information* 1(3),pp. 146–165.

Stereo. 2018. "Stats - OpenStreetMap Wiki." https://wiki.openstreetmap.org/wiki/Stats (March 17, 2018).

Vaz, Eric, and Jamal Jokar Arsanjani. 2015. "Crowdsourced Mapping of Land Use in Urban Dense Environments: An Assessment of Toronto." *Canadian Geographer* 59(2),pp. 246–255.

Zheng, Shudan, and Jianghua Zheng. 2014. "Assessing the Completeness and Positional Accuracy of OpenStreetMap in China." In *Lecture Notes in Geoinformation and Cartography*, pp.171–189.

Zielstra, Dennis, Hartwig Hochmair, Pascal Neis, and Francesco Tonini. 2014. "Areal Delineation of Home Regions from Contribution and Editing Patterns in OpenStreetMap." *ISPRS International Journal of Geo-Information* 3(4),pp. 1211–1233.

APPENDIX

The online version of questionnaire in this paper: https://www.wjx.cn/jq/7998309.aspx

The paper version of questionnaire in this paper:
1. Your gender:
A. male   B. female

2. Your age:
A. Less than 20   B. 20~29   C. 30~39
D. 40~49   E. 50 or over

3. Your educational background:
A. High school or less   B. Undergraduate
C. Master   D. Ph.D.

4. Your living City:
_____

5. Your occupation:
A. Geography or computer science related professionals
B. Company Staff   C. Student
D. Government staff   E. Freelancer
F. Other

6. Your monthly Income:
A. Less than 2000   B. 2000~3999

C. 4000~5999   D. 6000~7999
E. 8000~9999   F. 10000 or over

7. Have you heard about OSM?
A. Never heard about it   B. Heard but not used
D. Used   D. Used and contributed

8. Are you used to use electronic maps?
A. Yes   B. No

9. The frequently you use OSM:
A. Always   B. Occasionally
C. Almost not

10. **(Multiple choice)** The purpose you use OSM:
A. Find addresses, navigation needs
B. Work or learning needs
C. Personal hobby
D. Contribution of geography information
E. Other

11. **(Multiple choice)** Compared to other Web map, what are the advantages OSM have?
A. Open source, we can download the relevant geographic data
B. Geographic information detailed and accurate
C. Graphical interface looks better
D. Users can edit the map
E. Other

12. **(Multiple choice)** The inadequate of OSM in your eyes:
A. The OSM easily leaks out the national secrets and other sensitive information
B. The quality of OSM can't be guaranteed
C. The OSM hasn't adequate the geographic data
D. Other

13. Where is the main region you contributed in OSM?
A. Nearby of living and working place
B. Every new place in life
C. The region leak of OSM data
D. Others

14. **(Multiple choice)** What is your purpose when contributed OSM?
A. Doing contributions to the open source geographic data
B. Needs of job occupation
C. Hobby of own

15. How long time do you spend on contributing geographic information?
A. Less than 10 minutes   B. 10-19 minutes
C. 20-29 minutes   D. More than 30 minutes

16. Do you know OSM is an open source project?
A. Yes   B. No

17. Do you think highly of the concept of Open Source?
A. Very serious   B. Serious
C. Unserious   D. Unconcern

18. What is your experience about OSM after exploring the website?
A. Map quality is good   B. Similar to other map
C. Worse than other map   D. Have no time to explore
E. Other

19. Are you willing to use OSM after short presentation?
A. Very glad to use   B. Glad to try
C. Not sure   D. Not interest

20. **(Multiple choice)** Why don't you want to use OSM?
A. No need   B. Don't like electronic map
C. Be used to other electronic map   D. Other

21. Are you willing to spread OSM to others?
A. Very glad to   B. Glad to
C. Not sure   D. Not interest

22. What are your suggestions for this survey?
_____