International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:10, 2016

Analysis of Urban Population Using Twitter Distribution Data: Case Study of Makassar City, Indonesia

Yuyun Wabula, B. J. Dewancker

Abstract**—**In the past decade, the social networking app has been growing very rapidly. Geolocation data is one of the important features of social media that can attach the user's location coordinate in the real world. This paper proposes the use of geolocation data from the Twitter social media application to gain knowledge about urban dynamics, especially on human mobility behavior. This paper aims to explore the relation between geolocation Twitter with the existence of people in the urban area. Firstly, the study will analyze the spread of people in the particular area, within the city using Twitter social media data. Secondly, we then match and categorize the existing place based on the same individuals visiting. Then, we combine the Twitter data from the tracking result and the questionnaire data to catch the Twitter user profile. To do that, we used the distribution frequency analysis to learn the visitors' percentage. To validate the hypothesis, we compare it with the local population statistic data and land use mapping released by the city planning department of Makassar local government. The results show that there is the correlation between Twitter geolocation and questionnaire data. Thus, integration the Twitter data and survey data can reveal the profile of the social media users.

Keywords**—**Geolocation, Twitter, distribution analysis, human mobility.

I. Introduction

CURRENTLY, as the result of the urbanization, urban development has entered a new phase in which more than half of the world's population live in the urban area [1]. Increasing the number of population is one factor to ensure the sustainability of cities. The mobility of people is an integral part for understanding the city development as a whole [2]. Knowledge of urban human mobility can be a reference to the decision-maker and city planner to understand the urban structure, including urban planning and transportation [3]. Moreover, the characteristic of human movement from time to time is also different. Such as visiting shopping centers, parks, entertainment and other places, is an illustration of daily routines in urban areas that which influence human mobility.

Human mobility is a movement from one place to another destination [4]. In the literature, we found some approach to measuring the movement of urban citizens. Such information gathered through direct observation and a survey method to get information about how the people can interact with their surrounding [5]-[7]. They collected the information about the socio-economic characteristic of traveler, stability and their trip on a certain day. But on the other hand, this approach has some limitation such as the accuracy of the respondent to answer the questions and usually costly to implement, weakness to cover a large number of individuals and some problems of data validation. But along with the changing times, research also uses tracking technology devices to categorize urban mobility, such as mobile phone call tracking that describes the location where the call occurred [8], bank notes movement [9], subway smart card data [10], and GPS information of taxis [11]. However, mobility patterns and individual behavior within cities remains hard to understand, due to the lack a quantitative validation of the results, which's hard to obtain due to privacy concerns [12], [13].

Almost all smartphones now are equipped with the GPS feature. Through social media apps, people can share the virtual activity which can attach a location map to express happiness, pleasure, or an opinion about what they see, places they visit and where they are. With millions of data generated, social networking also provided a new opportunity for other fields to understand the city in the context of urban human mobility. By using the position coordinate of the user via cellular data in the form longitude and latitude coordinate [14], users can report their location and a particular venue anywhere in the world. In the literature, we found some research discussed the Foursquare social networking data to learn the social activity distribution of people [15] and place categories [16]. However, it was limited to showing information on specific users, a specific interest, and specific sites liked (hotels, restaurants, shopping, etc.). In the previous study, Gowalla and BrightKite social media data were used to investigate the human movement and seeks a new venue in around the cities [18], but this data cannot cover the whole area and only showed a specific location [17]. It indicates that social media data has been recognized by numerous research fields in recent years.

Despite the Twitter data having a large population, there are some open issues on a profile of Twitter users (example. age, gender, and occupation feature). Because the Twitter data cannot cover the whole range of user profiles, we then combined the real data from Twitter and questionnaire, to make the Twitter as an agent for understanding human mobility.

The objective of the research is to understand the movement pattern and distribution of the population in Makassar city based on a combination of Twitter data from the tracking results and the questionnaire data to perceive the Twitter user profile. There are two main focuses to build the research.

Yuyun Wabula is with the Faculty of Environmental Engineering, the University of Kitakyushu, 808-0135, Japan (phone: 090-8668-0584; e-mail: yuyunwabula@gmail.com)
B. J. Dewancker is with the Faculty of Environmental Engineering, the University of Kitakyushu, 808-0135, Japan (e-mail: bart@kitakyu-u.ac.jp).

Firstly, to analyze the urban human mobility pattern using Twitter social media data. Secondly, to define the land function based on the same individual visiting. Thus, the first aim of the study is to find a location or an area that people visit and then explain the land-use function. To validate these findings, we used the land use mapping released by the city planning department of the Makassar local government. In a further step, the second aim to investigate the relation between the Twitter data with the size of the population residing in these areas. As the validation, we used local data released by the statistic department of the Makassar city local government to compare with the Twitter data as shown in Fig. 1.

II. Method of Data Collection

There are two methods used to collect data:

A. Questionnaire Content

With regard to individual identity, the Twitter dataset did not cover the full identity of Twitter users. Thus, the questionnaire survey method was used to determine a broader user profile. The data gathered consists of personal information such as gender, age, education, address, as well as another a question related to urban mobility. For the purpose of this study, the questionnaire was posted online. A total of 200 valid questionnaires were obtained.

B. Twitter Dataset

On social media Twitter, individuals can post a short message is called "tweets", with up to 140 characters. To provide a personal status update, these days Twitter allows a variety of formats such as links to websites, and direct messages to other users [22]. One important feature of Twitter is that users can display a location map that reveals the time and place of where the status posted. Therefore, Twitter geolocation data will leave new traces for understanding human mobility. This research is focused in Makassar city, Indonesia which has an area of 17,577 km^2 and population density 8,010 km^2 with a population 1.409 million [23].

For data collection, we utilized the Twitter Streaming Application Program Interface (API). It is an application window that allows developers to access the program. It provides access for developers to read and write the Twitter data, e.g., record a new posting, read author profile, follower data, time zone and location information [20]. Our final Twitter data set consists of 40 days of data gathered from August 26 to October 4, 2015. It is constructed from Indonesian language tweets, which was filtered to reveal those tweets that contain the geolocation. In this study, we analyzed 268,353 data records from Twitter. In the next step, only data that contained user's location was processed.

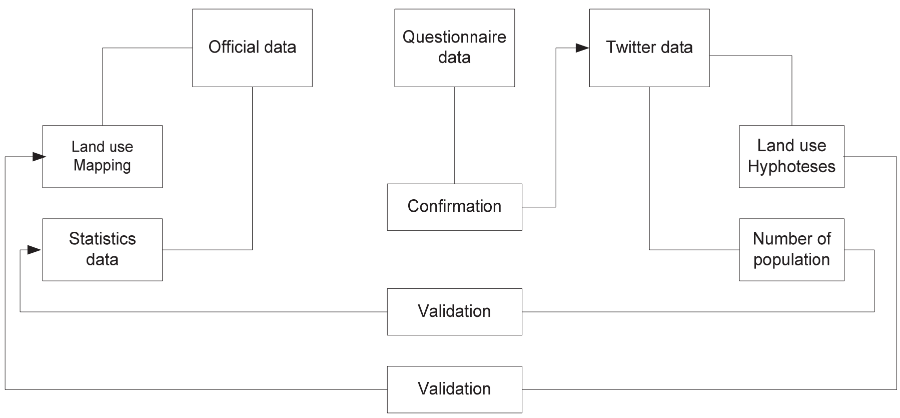

Fig. 1 System Flow

III. Data Analysis

The technique proposed for identification of the distribution of human mobility from the geolocation of a tweet is made up of two parts:

1) Collecting data through Twitter API was processed using Tableau Public 9.0. The goal is to observe the pattern of people movement and their distribution in a particular area. To learn for example when Twitter activity increases and decreases. The results of this analysis will produce a hypothesis about the existing land function. The analysis also will answer the question about tweet activities on weekdays and weekends.
2) Then, the data collected through questionnaire will be analyzed using Microsoft Excel and SPSS statistic. The results of this analysis will answer the question about the profile of users on the online social media service, Twitter.

IV. Results

A. Pattern of User Profile through the Questionnaire

The profile of the 200 respondents who completed the online questionnaire can be described as follows; regarding age, under 15 years old (one response) 1%, between 15-19 years old (19 replies) 10%, 20-29 years old (123 responses) 62%, 30-39 years old (46 replies) 23%, 40-49 years old (nine responses) 5%, and 50-60 years old, and over 60 years old, respectively (one answer) 1%. As for gender, the respondents consist of 34% male and 66% female. Regarding occupation, 36% were students, 16% teachers, 18% worked in a private company, 8%

in government, 3% were homemakers, and 13% described themselves as an entrepreneur, and 15% answered other. Table I shows the percentage profile of respondents.

TABLE I
RESPONDENTS BACKGROUND

Attributes		N = 200	%
Gender	Male	68	34%
	Female	132	66%
Age	<15	1	1%
	15-19	19	10%
	20-29	123	62%
	30-39	46	23%
	40-49	9	5%
	50-60	1	1%
	60>	1	1%
Occupation	Teacher	32	16%
	Private Company	35	18%
	Government	16	8%
	Student	72	36%
	Housewife	5	3%
	Entrepreneur	25	13%
	Other	15	8%

B. Pattern of User Activity

In a comparison of results from the survey regarding users Twitter activity, the visiting frequency revealed that most the Twitter activity was done at home with 169 responses (84.5%). This is then followed by workplace with 110 replies (55%), school/university with 82 answers (41%), the park with 34 responses (33%), mall and other shops with 66 replies (33%). Meanwhile, other locations include a coffee shop, the airport, and interesting places at 18 responses (9%). According to the time when the respondent is most active on Twitter, the morning recorded 49 respondents (25.4%), daytime - 50 respondents (25.9%), afternoon - 54 respondents (28%), evening - 119 respondents (61.7%) and anytime - 79 respondents (40.9%).

C. Pattern of Twitter User Profile through API

From the questionnaire results, analysis was conducted to learn the user Twitter profiles. Our initial assumption shows the age of users is in the range 15 years to 50 years olds, and the average is dominated by the 20 years to 39 years old age group. The gender category is 34% male and 66% female. For respondent occupation, this was mostly dominated by students, government employees, and private companies, respectively 36%, 26%, and 18%.

Fig. 3 demonstrates the visualization of users in different locations. Each coordinate location divided into 12 groups. Each group represents a district. The dots represent the number of check-ins, and the color indicates the different categories.

From the results of the grouping, we then offer a hypothesis about the number of Twitter user and the type of land function. To verify our assumptions, we compare the evaluation results to the official population data released by the department of statistics of Makassar City local government (see Table II). Then, to detect the land function, we used the official land use data published by the Department of Spatial Planning of Makassar City local government (see Fig. 3). We observed that 8% of Twitter users could be found in the Biringkanaya district, which has the largest population and largest area. And when compared to the questionnaire data, 14% of respondents reside in this district. In fact, most of the Twitter user activity took place around the airport, not in the residential area.

Then, for the Twitter users generated from the two districts; Bontoala and Tallo, respectively, 2% and 1%. According to the existing land function, the area is designated for warehousing. Our assumption that this area is not attractive to visit due to the lack of facilities and infrastructure in the surrounding environment to attract people. Twitter activity is only done around the home, which also affects the number of respondents, which was only 3%.

Another interesting finding is related to the Tamalandrea district, which has the highest number of respondents (18%) and Twitter users (11%), while the majority of them is the students. Based on the mapping activity (see Fig. 3), the brown color is more dominated by people activity in the university area. Then our hypothesis that land function is the education area. That is because the majority of students spend their daily routines in the home and the university. It can be proved by comparing the occupation type of respondents on questionnaire and the land function map released by the local government.

Meanwhile, 14% of Twitter activity came from the Tamalate district. When compared to the other regions, it was the second largest for Twitter distribution and the reason could be that this area, based on the land use data, is designated a tourism and culture area. There are two districts which represent the business area, Mariso 9% and Ujung Pandang 12%. Regarding the residential area, three districts have been identified; Panakukkang 20%, Rappocini 7%, and Manggala 7%.

V. CONCLUSION

This paper presents the use of geo-location Twitter as a way of understanding the movement of people. The results have shown that; firstly, the three largest districts for Twitter users in Makassar city, namely Panakukkang (20%), Tamalatte (14%) and the Ujungpandang district (12%). Secondly, based on our hypothesis, the largest distribution of human mobility is contained in the residential area. Thirdly, concerning the movement of people, we concluded that land function could also characterize the behavior of people. It can be proved that most of the respondents are students and connected to the Tamalandrea district, which is an education region. It means that there is a correlation between land function and people who inhabit a particular area associated with occupation. Fourth, questionnaire data can be used as a data source to determine the user profile of the social media service, Twitter. Because this study was limited to one city and only used geo-location data as the parameter, future research should compare the data with other cities and use of the content of social media, e.g., tags and comments to know what kinds of things they are talking about or whether they discuss related to urban issues?

TABLE II
COMPARISON OF EACH DATA SOURCE BETWEEN THE MAKASSAR CITY POPULATION [21], QUESTIONNAIRE, AND TWITTER

Districts	Wide Area Km²	Population N=1331760	Average	Questionnaire N=200	Average	Twitter N=253968	Average
Biringkanaya	48.22	190802	14%	27	14%	10266	8%
Bontoala	2.1	55910	4%	6	3%	4200	2%
Makassar	2.52	84014	6%	13	7%	9000	3%
Mamajang	2.25	60236	5%	7	4%	8400	3%
Manggala	24.14	127915	10%	20	10%	18000	7%
Mariso	1.82	56547	4%	5	3%	24322	9%
Panakukkang	17.05	56610	4%	25	13%	52580	20%
Rappocini	9.23	160499	12%	20	10%	19400	7%
Tallo	8.75	135216	10%	13	7%	2712	1%
Tamalandrea	31.86	109471	8%	35	18%	28720	11%
Tamalate	20.21	186921	14%	15	8%	35768	14%
Ujung Pandang	2.63	27141	2%	5	3%	33000	12%
Ujung Tanah	5.94	48531	4%	2	1%	2200	1%
Wajo	1.99	31947	2%	7	4%	5400	2%

APPENDIX

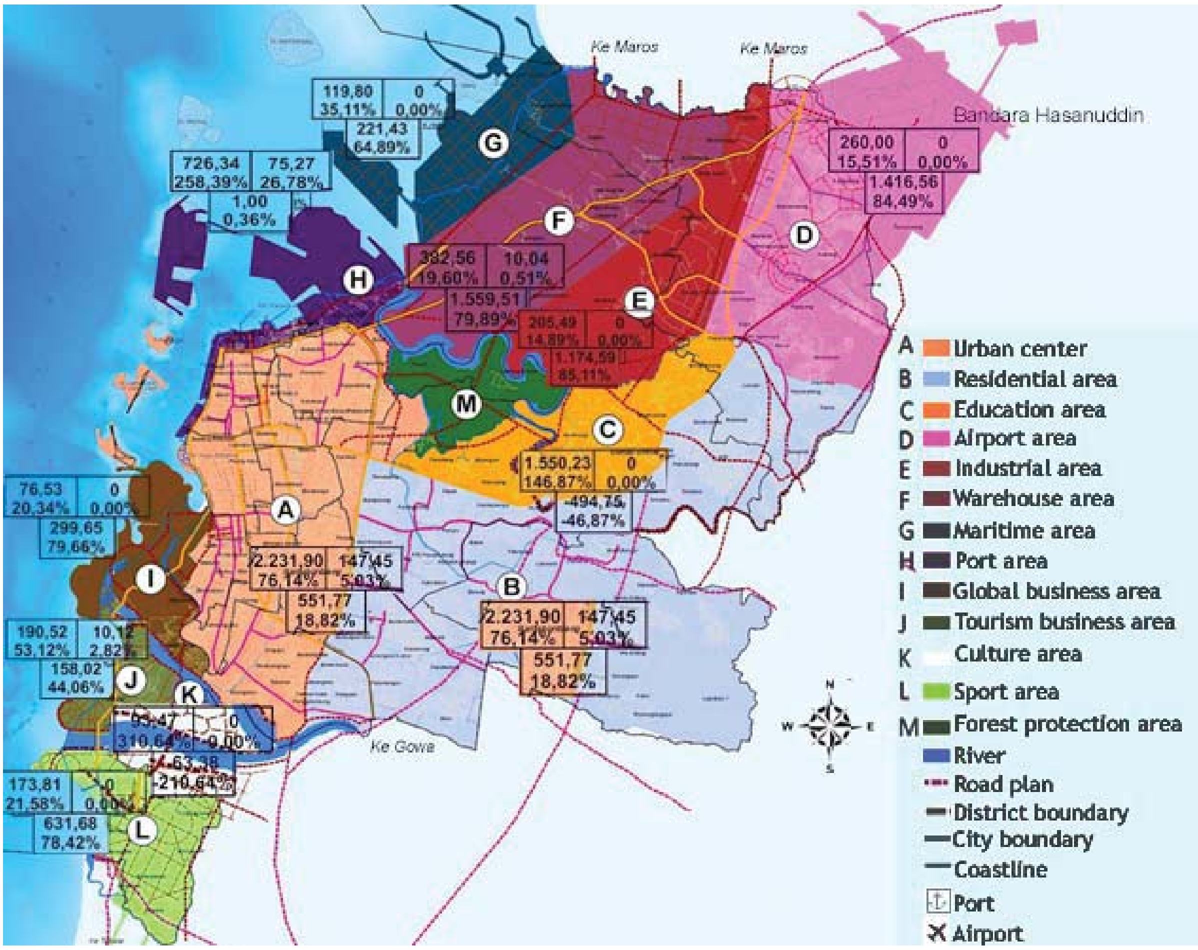

Fig. 2 Official land use maps of Makassar city government [19]

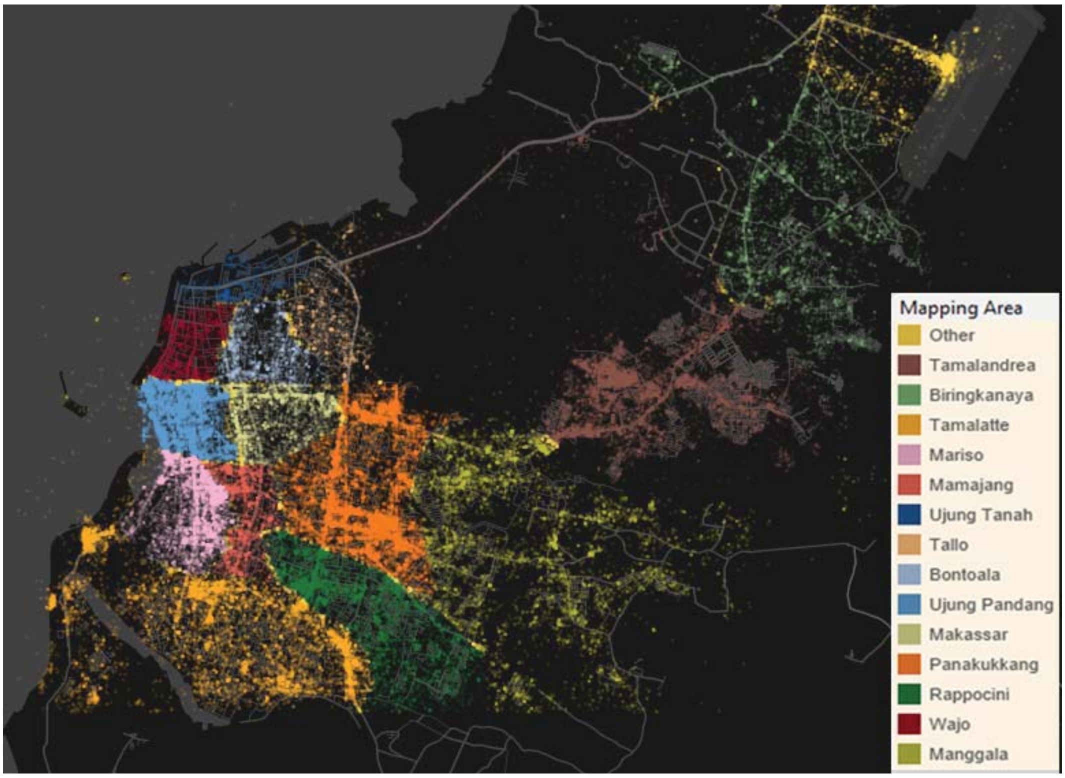

Fig. 3 Comparison the distribution of user check-ins for each district

ACKNOWLEDGMENT

This research was supported by the University of Kitakyushu, Directorate General Higher Education of Indonesia (DIKTI), and STMIK Handayani Makassar

REFERENCES

[1] Zheng, Y. U., Capra, L., Wolfson, O., & Yang, H. A. I., "Urban Computing: Concepts Methodologies, and Applications," 2014

[2] Wang, C., Taylor, J.E, "*Process Map for Urban-Human Mobility and Civil Infrastructure Data Collection Using Geosocial Networking Platforms*," 2015.

[3] Baedeker, S.B., Kist, C., Merforth. M., "Next Generation Urban Mobility Plans," 2014.

[4] An Overview: Land Use and Economic Development in Statewide Transportation Planning, A Report Prepared for the Federal Highway Administration by the Center for Urban Transportation Studies, *University of Wisconsin-Milwaukee, May*, 1999

[5] Ewing, R., Cervero, R., "A Synthesis. *Analyzing Public Policy*," pp.154–177. 2013.

[6] Hanson, S., "Perspectives on the geographic stability and mobility of people in cities," *Proceedings of the National Academy of Sciences of the United States of America*, *102*, pp. 15301–15306. 2005.

[7] Vilhelmson. B., "Daily mobility and the use of time for different activities. The case of Sweden," *GwJournal*, pp. 177-185, 1999.

[8] González, M. C., Hidalgo, C. A., & Barabási, A.-L., "Understanding individual human mobility patterns". *Nature*, *453*, pp. 779–782, 2008.

[9] Brockmann, D., Hufnagel, L., Geisel, T., "The scaling laws of human travel". *Nature*, 462, 2006.

[10] Hasan, S., Schneider, C. M., Ukkusuri, S. V., González, M. C., "Spatiotemporal Patterns of Urban Human Mobility," *Journal of Statistical Physics*, *151*, pp. 304–318, 2012.

[11] Jain, M. "A next-generation approach to the characterization of a non-model plant transcriptome," *Current Science*, *101*(11), pp. 1435–1439. 2011.

[12] Vazquez-Prokopec, G. M., Bisanzio, D., Stoddard, S. T., Paz-Soldan, V., Morrison, A. C., Elder, J. P., Kitron, U., "Using GPS Technology to Quantify Human Mobility, Dynamic Contacts and Infectious Disease Dynamics in a Resource-Poor Urban Environment," *PLoS ONE*, *8*(4), pp. 1–10, 2013.

[13] Frias-Martinez, V., & Frias-Martinez, E., "Spectral clustering for sensing urban land use using Twitter activity," Engineering Applications of Artificial Intelligence, 35, 237–245, 2014.

[14] Croitoru, A., Wayant, N., Crooks, A., Radzikowski, J., & Stefanidis, A., "Linking cyber and physical spaces through community detection and clustering in social media feeds," *Computers, Environment and Urban Systems*, *53*, pp.47–64, 2015.

[15] Noulas, A., Scellato, S., Mascolo, C., & Pontil, M., "Exploiting Semantic Annotations for Clustering Geographic Areas and Users in Location-based Social Networks," *The Social Mobile Web*, pp.32–35, 2011.

[16] Phithakkitnukoon, S., & Olivier, P., "Sensing Urban Social Geography Using Online Social Networking Data," *The Social Mobile Web*, pp.36–39, 2011.

[17] Noulas, A., Scellato, S., Lathia, N., & Mascolo, C., "A random walk around the city: New venue recommendation in location-based social networks, " *Proceedings - 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust and 2012 ASE/IEEE International Conference on Social Computing, SocialCom/PASSAT 2012*, pp. 144–153, 2012

[18] Cho, E., Myers, S. A., Leskovec, J., "Friendship and mobility," *17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1082. 2011.

[19] Spatial planning map of Makassar city 2010-2030. (2016, August 8). *http://darimakassar.com/rtrw-kota-makassar-2010-2030-2*

[20] Twitter. Open twitter streaming api. (2015, August 26). *https://dev.twitter.com/docs/streaming-api.*

[21] Central Bureau Statistic. Makassar in Number（2014）.

[22] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment," *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, pp. 178–185, 2007.

[23] Central Bureau Statistic. South Sulawesi in Number（2015）.