

## A method for studying the development pattern of urban commercial service facilities based on customer reviews from social media

Y.D. Wang <sup>a\*</sup>, B.T. Jiang <sup>a,b</sup>, X.Y. Ye <sup>c,\*</sup>

<sup>a</sup> State key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China - ydwan@whu.edu.cn (Y.W.)

<sup>b</sup> NingBo Planning and Geographic Information Center, 36 Heji Street, Ningbo, China - changbt@qq.com

<sup>c</sup> Department of Geography, Kent State University, Kent, USA - xye5@kent.edu (X.Y.)

\* Corresponding author. Email: ydwan@whu.edu.cn; xye5@kent.edu

### Theme Sessions, ThS 12

**KEY WORDS:** commercial service facility, analysis based on road network constraints, quantitative evaluation of urban planning, social network service

Urbanization is one of the most important human social activities in the 21st century (Chaolin et al., 2012). With an increasing number of people visiting cities, the provision of adequate urban service facilities, including public and commercial service facilities, in locations where people live has become an important guarantee of the success of urbanization. Exploring the commercial service facilities in a specific area of a city can help us understand the progress and trends of urban renewal in the area, provide a quantitative basis for evaluating the rationality of planning implementation, and facilitate an analysis of the effects of different factors on the regional development of a city (Schor et al. 2003).

In this paper, we proposed a data processing and analysis method for studying the distribution and development pattern of urban commercial facilities based on customer reviews. In addition, based on road network constraints, we explored the patterns contained in customer reviews data, including patterns for the spatial distribution and spatial-temporal evolution of facilities as well as the number of facilities and degree of satisfaction.

### 1. DATA AND STUDY AREAS

Sina-Weibo (e.g., <http://us.weibo.com>), a Twitter-like microblogging system, is the most popular microblogging service in China. Using web crawlers and Sina Weibo API, we collected 706,835 Sina-Weibo with 'Beijing rainstorm' as the keyword with time stamps between 00:00 AM on July 20 and 24:00 AM on August 10. Among collected, there were 26,050 Sina-Weibo texts which contained GPS information, and 10,988 of them were located in Beijing.

Dianping, a Yelp-like social media website, was founded in 2003 and is a leading urban life and consumption platform in China which providing third-party customer reviews. Dianping provides customers with information on local commercial services, including information on nearly all the local life service industries (e.g., food and beverage industry, movie industry, hotel industry, recreation and entertainment industry, beauty industry, etc.)

In the paper, the Dashangdi District of Beijing has been selected as the study object, and the information from customer reviews on the commercial service facilities within the Dashangdi District has been obtained from Dianping. Using

web crawlers, we obtained 35,394 customer review texts corresponding to 2,038 commercial service facilities of 13 different types in the Dashangdi District from May 1, 2004 to June 30, 2014.

### 2. CALCULATION OF THE DEGREE OF SATISFACTION FROM THE CUSTOMER REVIEW TEXTS

In our study, the numerical values of the degree of satisfaction, which range from 1 to 5 points, are used to measure how satisfied customers are with the commercial service facilities that they have reviewed. The essence of calculating the degree of satisfaction with a commercial service facility from its customer review is to classify the review text based on semantic meanings.

In this paper, the 35,394 customer review texts are classified into five types based on a topic model to quantify the degree of satisfaction with commercial service facilities. The process of calculating a degree of satisfaction from customer review texts consists of three steps: (a) constructing evaluation phrases for degree of satisfaction (terms) and extracting these terms from the customer review texts; (b) establishing a "document-term" model for the entire review text data set based on the terms and using the model to score the terms based on latent semantic indexing (LSI); and (c) determining the corresponding score of each review document based on the calculation result of the document topic.

### 3. SPATIAL DISTRIBUTION CHARACTERISTICS OF DIFFERENT TYPES OF COMMERCIAL FACILITIES

The characteristics of commercial service facilities require that these facilities be distributed on two sides of a road. By investigating the distribution densities of POIs on two sides of a road network, the degree of concentration of each type of commercial service facility on a road segment can be evaluated, and the prosperity level of each type of commercial service facility as well as the industrial agglomeration characteristics can then be determined. Therefore, when analyzing one type of commercial service facility, it is necessary to consider the road upon which this type of facility is located. Based on road network constraints, a method for determining the density of commercial service facilities distributed on a traffic network is

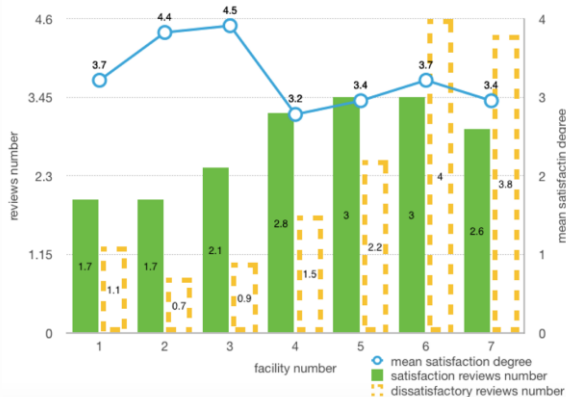
designed in the present study, and this method can be used to analyze the spatial distribution characteristics of commercial service facilities.

In order to analyse the concentration of commercial facilities along different segments of the entire road, the road network is first decomposed into linear units with a fixed length, 200m is selected in this study, and the shortest network distance between two POIs is used for measurement; thus, after a road is portioned into multiple segments with the same length, the POIs will be more accurately mapped onto a certain section of the road, which is advantageous for differentiating the densities along different segments of a road. After the number of POIs that are mapped to each road segment is calculated, the road segment within a bandwidth of  $r$  is selected as the neighbour of each individual road segment within the density range; then, the kernel density can be calculated using equation.

#### 4. ANALYSIS OF THE CHANGES IN THE DEGREE OF SATISFACTION WITH COMMERCIAL SERVICE FACILITIES

During the development processes of many service facilities, when one facility emerges and becomes popular, many similar service facilities will rapidly establish in the vicinity, and the number of and degree of satisfaction with this type of service facility over the entire road will increase and decrease with the intensifying competition. In this section, an analysis is performed of the correlation between the number of and degree of satisfaction with beauty facilities and the temporal-spatial changes in the degree of satisfaction.

As shown in Figure 1, we calculated the correlations among the number of facilities, mean degree of satisfaction, number of satisfactory customer reviews and number of unsatisfactory customer reviews.

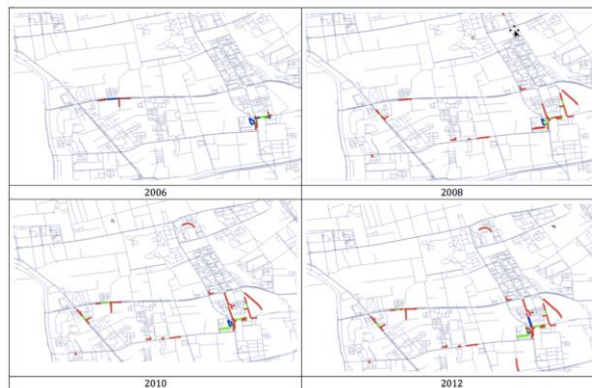


**Figure 1.** Relationship between the number of facilities and reviews

From the Figure 1, we can find the following conclusions. (a) For the distribution of commercial beauty service facilities in fierce competitions with one another, the correlation coefficient between the number of facilities and the mean degree of satisfaction in a single road segment is  $-0.5183$ ; thus, the two parameters are not highly linearly correlated with one another and there is no direct relationship between the mean degree of satisfaction and number of beauty facilities along a road segment. (b) When the number of facilities is 1~3, the correlation coefficient between the mean degree of satisfaction and number of facilities is  $0.9177$ ; thus, the mean degree of satisfaction is highly correlated to the number of facilities, and

a local distribution of 1~3 facilities can result in the regional maximization of the mean degree of satisfaction along a road segment.

Figure 2 shows the temporal-spatial changes in the beauty facilities from 2006 to 2012. In the Figure, unsatisfactory POIs (degree of satisfaction with a numerical value of 1~2.9 points) are marked in green, generally satisfactory POIs (degree of satisfaction with a numerical value of 3~3.9 points) are marked in red, and satisfactory POIs (degree of satisfaction with a numerical value of 4~5 points) are marked in blue.



**Figure 2.** Changes in the degree of satisfaction for beauty POIs

From the Figure 2, we can find that: (a) Over time, commercial beauty service facilities gradually expanded from artery roads to secondary roads and gradually formed an agglomeration in the Dashangdi District on the right; in addition, the central area on the map indicates a large number of commercial beauty service facilities agglomerated in large-scale department stores during this period, and this agglomeration created the fiercest competition in the beauty service field in this area. (b) From 2006 to 2012 the commercial beauty service facilities with a “general” degree of satisfaction were prevalent; however, the number of commercial beauty service facilities increased in the central area, which also concentrated the commercial beauty service facilities with satisfactory and unsatisfactory customer reviews, indicating that competition was extremely fierce in this area and suggesting a certain correlation between the number of facilities and degree of satisfaction.

#### REFERENCES

CHAOLIN, G. U., LIYA, W. U. and COOK, I. 2012. Progress in research on Chinese urbanization. *Frontiers of Architectural Research*, 1(2), 101-149.  
SCHOR, S. M., ARTES, R. and BOMFIM, V. C. 2003. Determinants of spatial distribution of street people in the city of São Paulo. *Urban Affairs Review*, 38(4), 592-602.