

# Social Benefit Evaluation of Transportation Policy Based on Social Media

Xin Wang, Xingqiang Zhang\*, Yonggang Song, Shan Zhu and Xueyuan Wang

MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Center of Cooperative Innovation for Beijing Metropolitan Transportation, Beijing Jiaotong University, Beijing, 100044, China

**Abstract**—Social Media data can truly reflect hot spot and attitude of people to transport policy. The scientific social media data mining and processing method can excavate effectively the public’s focus to transportation policy and provide a new transportation policy evaluation way. Firstly, the evaluation index system of transportation policy including its social benefit is established. Secondly, a Social Media Chinese text data mining scheme is presented, using Python and the information processing technology of Chinese synonym, fuzzy clustering weight of text data based on fuzzy clustering method. Then a social benefit evaluation index calibration method of transportation policy is presented. Finally, social benefit of Beijing public transportation pricing policy is evaluated, the public’s attention is transport comfort.

**Keywords**- intelligent transportation; parameter calibration; fuzzy clustering; social media; data mining; transportation policy component

## I. INTRODUCTION

New transportation policies are mainly to serve the public. The traditional traffic policy, evaluated based on data collected by government departments and related agencies, pays more attention to economic benefits. However Mattauch L[1] considered that the public’s behavior and attitude should be taken into transportation policy evaluation criterion. Study by Gal-Tzur[2,3] showed that traffic information and the public’s views on transportation policy could be investigated timely from Social Media and the investigated traffic information was used into transportation policy research. Based on Social Media data, Cao[4] analyzed traffic mood, and studied 2 policies, "breaking traffic yellow light" and "Chinese oil price". Krishen A. S.[5] thought that transportation policy framer could use Internet data to find what was the public’s interest. In conclusion, Social Media data could be used into transportation policy evaluation. So based on the public’s opinions, the evaluation index system of transportation policy including its social benefit was established. According to Social Media data, synonym Thesaurus method and text data fuzzy clustering method, a kind of weight calibration method of transportation policy social benefit evaluation indexes was established, and the method was used into case study of Beijing public transport price adjustment policy in 2014.

## II. TRANSPORTATION POLICY EVALUATION INDEX SYSTEM SELECTING A TEMPLATE

Traditional transportation policy evaluation index system

includes 3 aspects: policy influence, technical influence and benefit influence, in which the benefit influence should include economic and social benefit. The public’s opinion on transportation policy is a important components of social benefit. Comfort, economy and convenience are the main concerns of the public to transportation policy. The evaluation indexes of transportation policy, including social benefit indexes, are shown in Fig. 1.

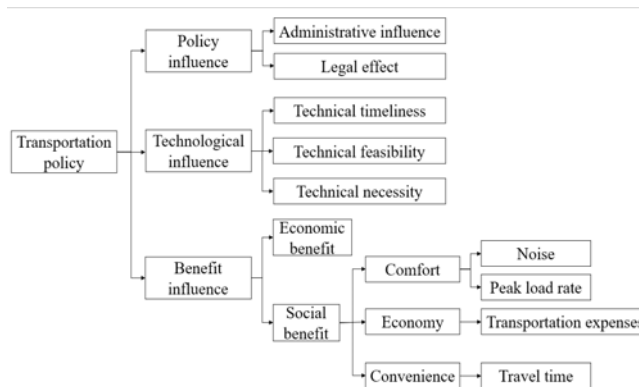


FIGURE I. TRANSPORTATION POLICY EVALUATION INDEX SYSTEM

Social Media, people communicate and express their wishes each other, is a public forum which is efficient, fair, true. How to establish a reasonable calibration method on evaluation indexes of the social benefit based on Social Media can truly reflect the public's evaluation on the new transportation policy, and obtain better the evaluation effect.

## III. CALIBRATION OF SOCIAL BENEFIT INDEX PARAMETERS

### A. Data Mining Scheme

1. Based on the Base64 and RSA2 encryption algorithm, using a Python procedure to login Social Media;
2. Downloading texts in the specified URL, and saving the texts format as id, name, content, timestamp;
3. Using the crawler script to login and download, and searching Social Media posters in Beijing with keywords form December 20, 2014 to January 31, 2015. For example, 3 887 posters with keyword “公交涨价”, 3 364 posters with keyword “调价” and 134 448 posters with keyword “地铁”.

**B. Synonym Segmentation Technique**

In order to mine deeply Social Media text data and improve their preciseness and accuracy, the synonym Thesaurus method, *HIT IR-Lab Tongyici Cilin (Extended)*.<sup>[6]</sup>, is used to analyze keyword semantic difference. There are 9 kinds of degree expression in Chinese words which is respectively indicative, overlapping, causal, comparative analogy exaggeration, exclamation. For example, using *HIT IR-Lab Tongyici Cilin (Extended)* to analyze Chinese “拥挤”, its classification code, EF03B01, and 22 synonyms, “挤”, “轧”, “蜂拥”, “挤挤插插”, “人山人海”, “水泄不通”, “磕头碰脑”, “前呼后拥”, “摩肩接踵” etc, are obtained.

Then a Python procedure is used to combine Chinese word frequency, and extract the Social Media posters containing these words. For example, The number of posts, containing keyword “拥挤”, is 1 781. After combining keyword “拥挤” and its synonyms, the number is 2 774. Social Media utilization increases by 55.76%..

**C. Text Data Fuzzy Clustering Method**

In the paper, FCM, direct distance fuzzy clustering method, is used to weight the Chinese synonyms based on its importance.

$$d(x_i, x_j) = |x_i - x_j| \tag{1}$$

$$r_{ij} = 1 - bd(x_i, x_j) \tag{2}$$

In which,  $x_i$  -degree weight of  $i$ th Chinese keyword.  $d(x_i, x_j)$  -difference degree of  $i$ th and  $j$ th keyword.  $r_{ij}$  -  $x_i$  and  $x_j$  similarity.  $b$  -a parameter. In the paper, supposing  $b=1$  and  $0 \leq r_{ij} \leq 1$ . The cluster matrix  $U$  and the similar set can be respectively expressed by Eq.3 and Eq.4.

$$U = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n1} & \dots & r_{nn} \end{pmatrix}_{n \times n} \tag{3}$$

$$[x_i]_U = \{x_j | r_{ij} = \lambda_q\} \tag{4}$$

In which,  $[x_i]_U$  -the similar set of the  $i$ th word.  $\lambda_q$  - similarity classification threshold. Classification rules: 1) if  $q=1$ , then  $\lambda_1 = 1$ ,  $[x_i]_U = \{x_j | r_{ij} = 1\}$ ,  $x_i$  and  $x_j$  are similar set. 2) if  $q=2,3,\dots$ , then  $\lambda_q < \lambda_{q+1}$ , element pair, similarity is  $\lambda_q$  determined in  $U$ , and the equivalence classification

corresponding to  $\lambda_q$  is obtained. 3) if  $q=Q$ , the classification is completed.

**IV. AHP METHOD BASED ON SOCIAL MEDIA**

A kind of AHP method, calibrating index weight based on Social Media poster number, is proposed. The method can fairly and scientifically reflect the evaluation index of traffic policy. The judgment matrix is expressed by Eq.(5).

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n1} & \dots & a_{nn} \end{pmatrix}_{n \times n} \tag{5}$$

$$a_{ij} > 0, a_{ji} = \frac{1}{a_{ij}}, a_{ii} = 1, (i, j = 1, 2, 3)$$

In which:  $a_{ij}$  is the influence level ratio of  $i$  and  $j$  criterion layer to the target layer, and is calibrated using 9 scales (Tab.1).

TABLE I. PROPORTIONAL SCALE VALUE PRINCIPLE

$a_{ij}$	Meaning
1	Influence of criterion layer $i$ is same as $j$
3	Influence of criterion layer $i$ is slightly stronger than $j$
5	Influence of criterion layer $i$ is stronger than $j$
7	Influence of criterion layer $i$ is greatly stronger than $j$
9	Influence of criterion layer $i$ is absolutely stronger than $j$

Firstly, the weighted keyword number is equivalent to people quantity concerning about transportation policy and attention degree. Then the equivalent Social Media poster number is used as the AHP criterion of 9 scales which is shown in Eq.(6) and Eq.(7):

$$\begin{cases} N_i = \sum_{k=1}^2 N_{ik} \\ N_{ik} = \sum_{j=1}^n \sum_{t=1}^m c_{jt} M_{jt} \end{cases} \tag{6}$$

$$\begin{cases} N_i : N_j \approx 1, a_{ij} = 1 \\ N_i : N_j \in (1 + 0.5(r-1), 1 + 0.5r], a_{ij} = 1+r \\ N_i : N_j > 5, a_{ij} = 9 \end{cases} \tag{7}$$

In which,  $k$  -public satisfaction status,  $k=1$  satisfied,  $k=2$  dissatisfied.  $N_i$  -weighted Social Media poster number of  $i$ th criteria layer.  $N_{ik}$  -weighted Social Media poster number of  $i$ th criteria layer when the public satisfaction is  $k$ .  $M_{jt}$  - $t$ th keyword number of  $j$ th index.  $c_{jt}$  - the  $t$ th keyword fuzzy



TABLE IV CONTINUE

昂贵	$c_1 = 2$	高昂、昂贵
	$c_2 = 1$	值钱、贵
缓慢	$c_1 = 2$	磨磨蹭蹭、慢慢悠悠、慢条斯理、慢悠悠
	$c_2 = 1.5$	悠悠、慢吞吞、放缓、迟迟、款款、徐徐、迟缓、冉冉、缓缓、暂缓
	$c_3 = 1$	缓、缓慢、磨蹭、减缓、慢性、舒缓、慢
便宜	$c_1 = 2$	最低价、物美价廉、价廉物美
	$c_2 = 1.5$	公道、低廉、价廉、廉价、贱
	$c_3 = 1$	便宜、廉、低价

Chinese synonym numbers were merged based on the compiled Python program. The clustered and weighted keyword numbers were shown in Fig. 2.

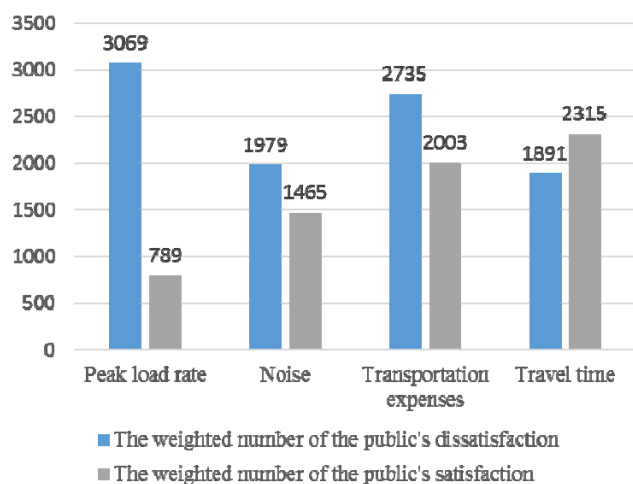


FIGURE II. WEIGHTED KEYWORD NUMBERS OF SECONDARY LEVEL INDEXES

The first level index weights of social benefit were calibrated using the AHP method. Based on the Social Media poster number comparison, the ratio between travelling comfort and economy was 1.48, the ratio between travelling economy and convenience was 1.25, the ratio between travelling comfort and convenience is 1.85. Then the AHP judgment matrix A was:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 1/2 & 1 & 2 \\ 1/3 & 1/2 & 1 \end{pmatrix}$$

The maximum eigenvalue  $\lambda_{\max} \approx 3.0092$ , the feature vectors  $W = (0.539, 0.297, 0.164)^T$ , the random consistency

index  $CI = \frac{\lambda_{\max} - n}{n - 1} = 0.0046$ ,  $RI = 0.58$ , the consistency

ratio index:  $CR = \frac{CI}{RI} = 0.0079 < 0.1$ .

Based on the second layer index weights, the first layer index weights, the positive and negative Social Media poster numbers of the second layer indexes, the public support quantity and opposition quantity was respectively 2189 and 3843, which showed the public did not support the price adjustment policy. And the public were more concerned about whether the travelling comfort was improved.

## VI. CONCLUSIONS

According to the Python web crawler script, the extended Chinese synonym information processing technology and the fuzzy clustering method, the Social Media transportation policy text data were weighted in which the weighted keyword numbers were equivalent to the people quantity and concern degree. Based on the equivalent Social Media poster numbers, a kind of 9-scale AHP index calibrated method of traffic policy social benefit was established. The new travelling ticket policy of Beijing public transport on December 28 2014 was used as case study, the results showed that the public did not support the new policy and paid most attention to the travelling comfort.

## REFERENCES

- [1] Mattauch L, Ridgway M, Creutzig F. Happy or liberal? Making sense of behavior in transport policy design[J]. Transportation Research Part D Transport & Environment, 2015, 45(S1):64-83.
- [2] Gal-Tzur A, Grant-Muller S M, Kuflik T, et al. The potential of social media in delivering transport policy goals[J]. Transport Policy, 2014, 32(32):115-123.
- [3] Gal-Tzur A, Grant-Muller S M, Minkov E, et al. The Impact of Social Media Usage on Transport Policy: Issues, Challenges and Recommendations[J]. Procedia - Social and Behavioral Sciences, 2014, 111:937-946.
- [4] Cao J, Zeng K, Wang H, et al. Web-Based Traffic Sentiment Analysis: Methods and Applications[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(2):844-853.
- [5] Krishen A S, Khan A. Interpretation of Public Feedback to Transportation Policy: A Qualitative Perspective[J]. Transportation Journal, 2014, 53(1):26-43.
- [6] ZHANG G D, ZHANG H X. Classification Algorithm Based on Semantics and Text Feature Weighting[J]. Application Research of Computers, 2012, 29(12):4476-4478.