

Multifunctional Product Marketing Using Social Media Based on the Variable-Scale Clustering

Ai WANG, Xuedong GAO

Abstract: Customers' demands have become more dynamic and complicated owing to the functional diversity and lifecycle reduction of products which pushes enterprises to identify the real-time needs of distinct customers in a superior way. Meanwhile, social media turned as an emerging channel where customers often spontaneously can express their perceptions and thoughts about products promptly. This paper examines the customer satisfaction identification and improvement problem based on social media mining. First, we proposed the public opinion sensitivity index (POSI) to uncover target customers from extensive short-textual reviews. Subsequently, we presented a customer segmentation approach based on the sentiment analysis and the variable-scale clustering (VSC). The approach is able to get several customer clusters with the same satisfaction level where customers belonging to each cluster have similar interests. Finally, customer-centered marketing strategies and customer difference marketing campaigns are planned under the shadow of customer segmentation results. The experiments illustrate that our proposed method can support marketing decision marketing in practice that enriches the intention of the current customer relationship management.

Keywords: customer satisfaction; scale transformation; short text clustering; social media mining

1 INTRODUCTION

Political factors have been increasing the uncertainty of our fragile global economy recently, and unpredictable trade risks might last even longer [1, 2]. Enterprises, especially multinational firms, must focus on the continuous improvement of their product quality and service level to survive [3]. Thus, the key issue depends on whether the enterprise is able to understand and respond to the customer needs.

However, customer demands have gradually become more dynamic and complicated due to the functional diversity and lifecycle reduction of products [4]. That pushes enterprises urgently to mine the real-time needs or opinions of different customers through appropriate techniques and methods. Although the traditional customer satisfaction investigation (usually based on questionnaires) is able to get the reliable customer attitude, relatively long data acquisition time and high communication cost make it hard to meet today's business requirements [5, 6].

Contributing to the rapid development of information network technology, social media comes into being and shows excellent advantages on the social interaction and information exchange [7]. Since the convenience and freedom of social media, it turns into an emerging channel where customers often spontaneously express their feelings and ideas about products [8,9]. Hence, social media platforms unconsciously provide an open data source, which stores plenty of real-time customer opinions on various products of different companies even from their competitors [4, 10-12].

Several researchers have noticed the potential value of social media, and previous works are organized mainly from two perspectives. As for the service-oriented enterprises, the study related to improving the business performance through mining customer-concerned factors gain more popularity. Zhao et al., [13] and Guo et al., [14] study the customer satisfaction influence factors gathering from online reviews to improve the management performance of hotels. One important conclusion illustrates that online reviews positively influence the

electronic word of mouth, and thus positively influence customers' overall satisfaction. Geetha et al., [15] confirm the consistency between the online ratings and actual customer feelings across two different categories of hotels. The research also suggests that managers of budget hotels should pay more attention to staff performance, compared with other premium hotels.

As for the product-oriented enterprises, the study related to direct the product planning through mining customer-generated topics gains more popularity. Jeong et al., [4] propose an opportunity mining approach based on social media modelling for product planning. According to the topic discovery and sentiment analysis method, the research evaluates the satisfaction level of each product topic and devises the development strategies referring to the primary keywords of topics. Jan [16] presents an automatic opinion mining method of customers' textual online reviews. The algorithm provides the word cloud of each product, which could enlighten enterprises on the production operation in the next stage.

In a word, researches above mostly apply the social media analysis to improve the enterprise competitiveness of the core business and then affect customers' feelings or experience. However, incentive strategies and marketing activities could directly promote customer satisfaction if they can arouse customers' interest accurately. Therefore, this paper studies the problem of customer satisfaction identification and improvement based on social media mining methods, which enriches the intention of the current customer relationship management (CRM).

The main contributions are as follows. First, we propose the public opinion sensitivity index (POSI) to measure the importance and popularity of a customer review. The POSI could be utilized to find target customers from extensive initial textual reviews. Second, we present a customer segmentation approach based on the sentiment analysis and the variable-scale clustering (VSC). The approach is able to get several customer clusters with the same satisfaction level, and also customers belonging to each cluster have similar interests or concerns. Third, specific customer-centered strategies are provided to

support the enterprise marketing decision following the results of our customer segmentation approach.

The rest of the paper is organized as follows. Section 2 presents previous works related to this research, including the customer satisfaction analysis and social media mining. Section 3 describes our main methodology together with related data analysis tools in detail. We present our experiment analysis that verifies the effectiveness of the proposed approach in Section 4. And the paper is concluded in Section 5.

2 LITERATURE REVIEW

2.1 Customer Satisfaction Analysis

Customer satisfaction refers to the difference between customers' expected effect and their actual feeling after purchasing the product or receiving the service [16]. Facing the fierce market competition, high customer satisfaction will promote enterprises to win more customer loyalty and earn more customer value. On the contrary, low customer satisfaction will cause much customer loss and make enterprises hard to maintain their competitiveness. Therefore, customer satisfaction is one of the best indicators reflecting the future profitability of enterprises.

Since different enterprises operate characteristic products and service, they have to establish their own customer satisfaction indices (CSI). For example, Helia et al., [17] calculate the CSI for a health referral centre from five dimensions that is reliability, responsiveness, assurance, empathy, and tangibility. Masound [18] builds a five-layered CSI system for e-commerce enterprises, consisting of the product layer, service layer, information technology, and network systems layer, payment system layer, and finally the loyalty and bonus program layer. Take the product layer as an example, and it is divided into six aspects, i.e., product customization, product price, product information, product scope, product quality, and product warranty indices. Besides, several countries have developed their standard CSI, such as American (ACSI), Swedish (SCSI), and Korean (KCSI) [19-21].

According to the CSI, enterprises attempt various ways to gather the basic information. Common approaches depend on establishing the information feedback and consultation system, as well as conducting the satisfaction questionnaire survey [16]. The customer complaint system facilitates the direct dialogue with customers via the telephone contact or on-site visit, which enables enterprises to know the satisfaction degree of different customers. Similarly, the statistical analysis on a large number of satisfaction questionnaires also successfully quantifies the customer satisfaction degree. However, these approaches have the same disadvantages, i.e. the extended data acquisition time and high communication cost.

In this regard, the network information technique provides new methods for customer satisfaction identification. On the one hand, a great many online platforms, like e-commerce websites, have already opened the rating function, which invites customers to score the product or service they offered, and the customer satisfaction degree could be derived from the evaluation scores. Ding et al., [22, 23] estimate the customer satisfaction of the personalized cloud service by

establishing a satisfaction function. They figure out that the strength of customer satisfaction is slightly increased when the actual feeling surpasses a certain value (expectation), while it is significantly reduced when the actual feeling falls below expectation. Let T_{ui} represent the actual feeling (score) of customer C_u after purchasing product P_i or receiving service S_i offered by an enterprise, and T^{exp} represents the expectation of customer C_u , the customer satisfaction function C_u is defined as [22]:

$$C_{ui} = \begin{cases} T_{ui} T_{ui} \geq T^{exp} \\ T_{ui} (T_{ui} - T^{exp} + 1)^\delta T_{ui} \leq T^{exp} \end{cases} \quad (1)$$

where the parameter $\delta \in N^+$ reflects the customer's tolerance to low-quality P_i or S_i .

Although customer satisfaction can be directly calculated via Eq. (1), enterprises still encounter the challenge of missing value, since most of the customers ignore the evaluation process due to their carelessness or unwillingness.

On the other hand, customers would like to post online reviews to express their attitudes towards some products or services on public social media platforms, which implies their satisfaction degree. The discussions usually consist of textual words, emoticons, and even pictures, etc. Jeong et al., [4] gather the online reviewing topics of a mobile phone product and then calculate the sentiment stock of every topic based on deep learning methods. The sentiment polarity of a topic is determined by its keyword weight (frequency) and the sentiment intensity of each keyword. Besides, they transform the sentiment stock of product topics to a scale of 0-10, that is satisfaction level, which successfully supports the subsequent application. Hence, this research selects the social network mining method to study the customer satisfaction identification of multifunctional products.

2.2 Social Media Mining

Social media mining aims to discover the inherent information and knowledge using the online data of social media platforms, which requires high data processing techniques including social media data gathering, pre-processing, modelling and analysis (see Fig. 1).

Web crawling is the most common technique to gather numerous online reviews in the text or picture format on social media [24]. Although crawlers can scrape the target dataset conveniently and continuously from any websites, they consume resources of visited systems and will cause the load and schedule issue [25]. To improve the data extraction efficiency, several crawling tools could be installed and utilized directly, such as Scrapy (scrapy.org), python-requests (docs.python-requests.org), and bazhuayu (www.bazhuayu.com).

Different from other websites, the web crawling results of social media platforms always exist in the short content, big noise, low normativity and high sparsity feature, which challenges the traditional text mining techniques. [26]. Certain data preparing and modelling methods should be designed especially for social media mining.

Traditional data pre-processing work of text mining mainly contains the word segmentation and stop words

removal. If we implement the word segmentation directly on the initial short-textual data of social media, it is bound to drive a high dimensional sparse model with massive noise in the next text modelling stage. Thus, appropriate text filtering work is of great significance utilizing the extra descriptive statistics information provided by social media platforms. For instance, Howells et al., [27] indicates that online users usually express their support of a post on Facebook by a click of "like," which enables us to filter valuable short texts according to the number of "likes" that a post gets.

There are two classic approaches of the text representation, that is the vector space model (VSM) and word embedding [28]. The VSM maps a document to a great many content-related words or phrases, which successfully translates the textual document calculation to the vector calculation [29]. Word embedding is widely utilized in the natural language processing (NLP) that contains the one-hot representation and distributed representation [32].

After solving the data pre-processing and modelling problem for the unstructured short texts of social media, various common text mining applications could be utilized directly to the social media analysis, e.g., social media topic modelling, sentiment analysis, short text classification, etc. Notably, the sentiment analysis that focuses on discovering people's polarity of sentiment implied in textual words [38] could form the basis of our customer segmentation method by estimating the satisfaction degree of every online customer.

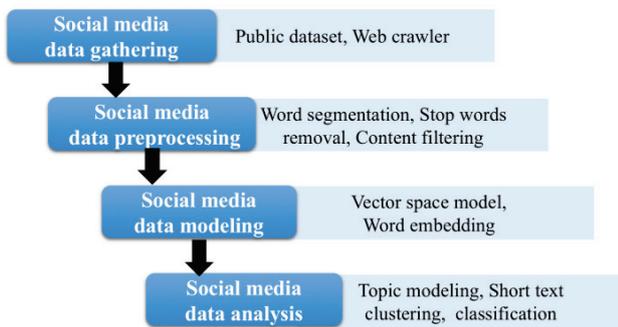


Figure 1 The research framework of social media mining

3 RESEARCH METHOD

In this section, we present a customer satisfaction estimation method using social media data. There are three significant issues that need to be solved. First, filter and exclude irrelevant and worthless reviews or metadata collected from social media platforms, such as advertisements, news, etc. Second, identify the essential and active customers regarding the numerous textual reviews, who are more probably to become the target audience of enterprises' marketing campaigns. Third, divide the customer base and reveal their demographics, perceptions, and concerns, to better support marketing decision making. Relative solutions are proposed in the following three phases.

3.1 Data Gathering and Preprocessing

In the beginning, we decide a target product and collect all the related online data from the social media platform via an open source crawling tool. The retrieval content not only contains all the textual reviews that at least mentioned the target product once but also includes the descriptive information of every review (e.g., number of forwarding, number of comments, number of likes), as well as the necessary information of every social media user.

According to the characteristic of the social network, the more significant a review is, the more it spreads, and vice versa [39]. We assume that social media users only interact with valuable and attractive reviews. Therefore, the social media data cleaning is achieved by setting coefficients on the descriptive information of the whole reviews.

3.2 Target Customers Identification

After data pre-processing, we start the formal word segmentation and stop words removal, where multiple reviews from the same user are merged for the completeness. Although marketers are eager to know the personality of consumers as much as possible (that is not limited to their past experiences, brand loyalty, perception of brands) when planning a marketing campaign, what they care about most is their response to campaigns, which might rapidly form the public opinion of an enterprise in a short term. Thus, we propose the public opinion sensitivity index (*POSI*) to measure the influence of online review(s) by the same customer to the enterprise's current public opinion.

Given the keywords of the enterprise's current public opinion on the target product, the public opinion sensitivity index (*POSI*) is expressed as:

$$\mu_i = \lg \left(\sum_{k=1}^N tf_{ik} * H_k \right) \tag{2}$$

where N is the total number of keywords in the public opinion, $H_k \in (0,1]$ is the importance of keyword w_k and $\sum_{k=1}^N H_k = 1$, tf_{ik} is the term frequency of word w_k in the reviews of customer C_i .

Customers with higher *POSI* either have larger frequency or higher importance, both represent that they are exactly the active and important target audience of enterprise's marketing campaigns.

3.3 Customer Segmentation

Although the *POSI* is able to distinguish target customers in terms of their short-textual reviews, it is still unable to satisfy marketers when creating deep insight into the consuming markets. Therefore, a customer segmentation algorithm based on the VSC is proposed to provide rich customer information for marketers in detail (see Fig. 2).

The algorithm consists of three steps as (1) Classify customer opinions through the sentiment analysis. At this point, customers have been classified into different groups following their sentiment attitudes (i.e., positive, neutral,

and negative), which greatly contributes to developing marketing strategies. (2) Segment customer bases via the VSC. Since multiple reviews of one customer have been merged during data preprocessing, customers with similar interests or concerns are divided into the same customer base according to the short text clustering results. (3) Identify customer characteristics by the VSC. Identify the key characteristic of every customer base to support enterprises planning specific marketing campaigns.

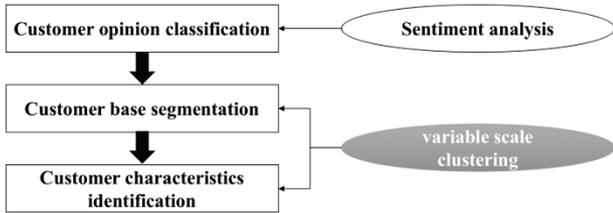


Figure 2 The process of the customer segmentation approach

As is known that people naturally are able to analyze and decide a problem from different perspectives, hierarchies, and dimensions spontaneously, that is referred to as scale transformation (ST) [42]. According to the basic definitions (i.e., the concept space (CS), the scale transformation rate (STR), and the granular deviation (GrD)) in [41], we propose a variable-scale clustering (VSC) method to solve the customer segmentation problem.

Given a dataset D , the CS of attributes in D , the scale transformation threshold S_0 , and the initial clustering parameter k , the pseudo code of the VSC is shown in Algorithm 1.

Algorithm 1 *AttributeScaleUpClustering*(D, CS, S_0, k)

```

1:  $C = D.initialCluster(k)$ 
2:  $R_0 = \max(GrD(C_i.qualified))$ 
3:  $D.delete(C_i.qualified)$ 
4: for  $D \neq \emptyset$  do
5:   for all  $A_j \in D$  do
6:     if  $STR(A_j, CH(A_j)) < S_0$  then
7:        $D.update(A_j, CH(A_j))$ 
8:     break
9:   end if
10: end for
11:  $R_0 = \max(R_0, GrD(C_i.closestQualified))$ 

```

```

12:  $C = D.Cluster(k-count(C_i.qualified))$ 
13: for all  $C_i \in C$  do
14:   if  $GrD(C_i) < R_0$  then
15:      $D.delete(C_i)$ 
16:   end if
17: end for
18: end for

```

The VSC overcomes the problem that the characteristics of clusters are less prominent especially handling the high dimensional dataset, of the traditional clustering algorithms. The time complexity of VSC is $O(nmkt)$, whereas the number of instances, m is the number of attributes, k is the number of clusters, and t is the number of iterations.

4 RESULTS AND DISCUSSIONS

Wechat, QQ, and Weibo are the three most popular social media networks in China and their usage from Dec 2017 to Jun 2018 is shown in Fig. 3. Currently, the usage rate of Wechat and QQ remains relatively stable, while Weibo gains a substantial increase due to the prevalence of short videos and multi-channel network (MCN) institutions [39]. Hence, this paper takes Weibo as the data source of social media, and the data gathering scheme is shown in Tab. 1.

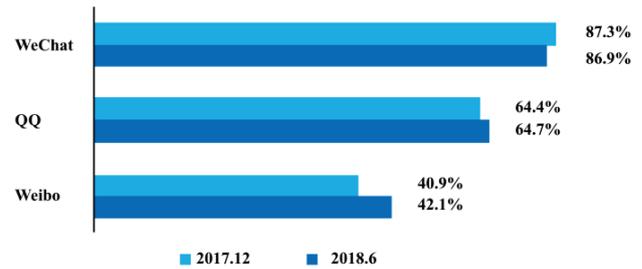


Figure 3 The usage rate of major social media in China from Dec 2017 to Jun 2018

We apply the professional web crawling tool, Scrapy, to collect approved online reviews on Weibo, and all the experiments are performed in OS X (10.11.3) environment on a machine with 8GB RAM.

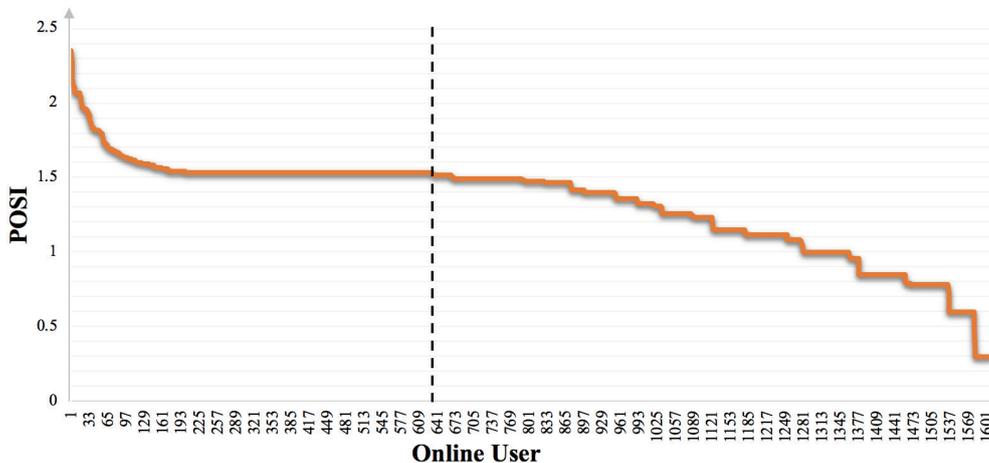


Figure 4 The POSI of qualified social media users

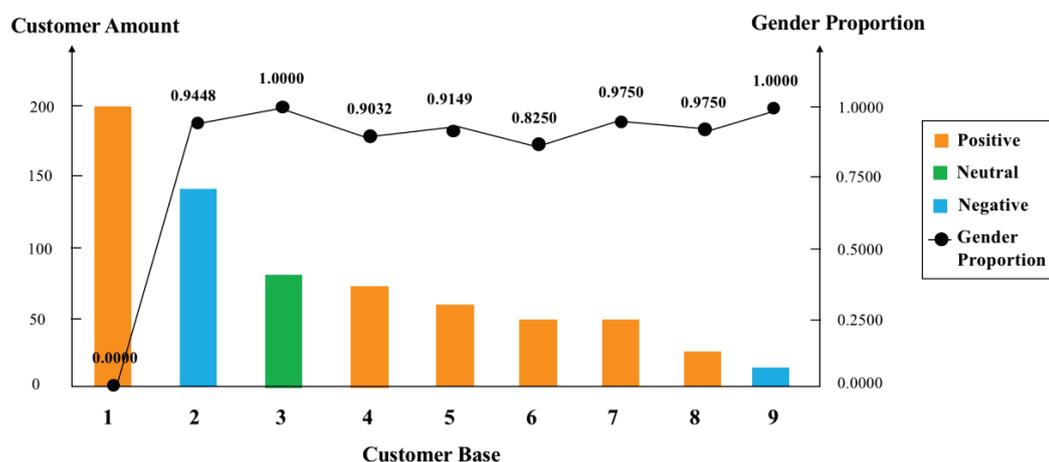


Figure 5 The customer segmentation results based on the VSC

Table 1 The data gathering scheme

Social Media	Weibo		
Target Product	iPhone X		
Time Period	2018-06-18 00:00to 2018-06-24 24:00		
Retrieval Conditions	(1)Original microblogs (2)Social media users have passed the personal identity authentication		
Data Structure	Object	Descriptive Information	Customer Information
	Customer reviews	Number of forwarding, Number of comments, Number of likes, Release time	Gender, Birthday, Country, Education, Affiliation, Occupation, Position

Table 2 The statistic results of the VSC

No.	Customer Satisfaction	Proportion	Avg. Age	Interests / Concerns
1	Positive	0,0000	22,8(0,3046)	After-sales service, promo
2	Negative	0,9448	24,9 (0,3379)	Video apps
3	Neutral	1,0000	24,7 (0,3034)	Sales, consumer
4	Positive	0,9032	23,6 (0,2742)	Running
5	Positive	0,9149	26,1 (0,3404)	Movie
6	Positive	0,8250	26,3 (0,3750)	Discount, product, gifts
7	Positive	0,9750	23,8 (0,4000)	Celebrity
8	Positive	0,9500	24,6 (0,2500)	Photograph
9	Negative	1,0000	24,1 (0,3671)	Member, World Cup

4.1 Experimental Results Analysis

Following the requirements of data gathering in Tab.1, there are 15165 original microblogs related to iPhone X during the valid period, while identity authorized online users posted only 4350 of them. Hence, we take 4350 microblogs as the initial social media dataset of our experiments.

The descriptive information is applied to the data preprocessing, where we set the minimum likes number as 3, the minimum comments number as 2, the minimum forwarding number as 1. Consequently, we gain 1607 qualified microblogs in total. It can be seen that all the irrelevant and worthless microblogs are excluded, which is consistent with the investigation result that over 40% reviews on the social media are advertisements, news, or other spam messages [40]. We merge multiple microblogs posted by the same user, and finish the word segmentation and stop words removal through professional tools in Section 2.

Then, we begin to identify the target customers from qualified social media users via the POSI (see Fig. 4). It can be seen that the POSI curve stays in a rapid decline trend in general, while part of it is relatively gentle. In order to identify the online users that truly affect the current

public opinion of iPhone X, we finally choose the top 634 users as our target customers.

Fig. 5 presents the customer segmentation results based on the proposed method VSC. Different colour represents different customer satisfaction (sentiment), i.e., positive, neutral, and negative, and each rectangle represents a customer base distinguishing the gender factor. We find out that (1) the distribution of the customer structure is extremely unbalanced, and the first two customer bases own over 54% customers. (2) there are nine customer bases in total, and five in positive, two in negative, one in neutral, which means most of the customers are currently satisfied with iPhone X. (3) The male proportion of customer base 2-9 is quite high (over 90%) and two of them (that is customer base 3 and 9) even reach 100%. However, the largest customer base has only the female.

What's more, Tab. 2 also provides further insight into the customer segmentation results, including customer satisfaction, gender proportion, average age, and customer characteristics (interests or concerns) of every customer base. The number in brackets of the attribute Avg. Age represents the age missing rate of each customer base and floats about 30%. And most importantly, the VSC discovers the various customer characteristics known as interests or concerns being hidden in customers' reviews.

These characteristics not only represent the similarity of customers but also are at the lowest (detailed) conceptual hierarchy, which further verifies the effectiveness and feasibility of the VSC in practice.

Finally, marketers could make reasonable marketing strategies on behalf of the estimated customer satisfaction and characteristics (see Section 4.2).

4.2 Customer-Centered Marketing Strategies

This section illustrates how to make marketing decisions using the results of the VSC. Firstly, decide customer-centered marketing strategies through customer satisfaction. For example, there are totally three different satisfaction levels in Tab. 2, i.e., positive, neutral, and negative. We could offer the precise benefits to positive customers, improve the stickiness of neutral customers, and increase the feedback collection of negative customers respectively.

Secondly, plan customer personalization marketing campaigns following the proposed marketing strategies. As for the precise benefits strategy, customer bases 1, 5, 8 most probably respond to the exclusive benefits of their interested software or apps. Customer base 7 might look forward to the intangible spiritual enjoyment (like meeting favourite celebrities), while customer base 6 might expect the tangible material reward (like valuable or saleable gifts). And termly pushing the running related information, such as distinctive or available marathon races to customer base 1 could improve their satisfaction. Referring to these, various marketing campaigns for positive customers could be designed. As for the customer stickiness and feedback collection improvement strategy, marketers could make full use of the known customer concerns to set up a more effective conversation and gain better solutions.

Last but not least, evaluate all the possible marketing plans and make the final decision. Since experience marketers are likely to brainstorm lots of marketing plans towards different customer base, it is necessary to establish an evaluation measurement of marketing plans. Let P_{ij} represents the j^{th} marketing plan of customer base CB_i , the evaluation indicator of P_{ij} as:

$$\text{Eval}(P_{ij}) = \partial_i * \frac{POSI_i}{c_{ij}} \quad (3)$$

where $c_{ij} \in (0,1)$ is the normalized cost of P_{ij} , ∂_i is the of CB_i , and $POSI_i$ is the average public option sensitivity of CB_i .

The marketing plans with higher evaluation score would earn larger return rate, compared to the lower ones. Hence, marketers could easily select the candidate campaigns following the descending order of evaluation scores under funding constraints.

5 CONCLUSIONS

The fierce market competition pushes enterprises to seek more effective and flexible techniques for continuously keeping the competitive advantages. Therefore, this paper focuses on the problem of customer satisfaction identification and improvement on the

perspective of social media mining. The public opinion sensitivity index ($POSI$) was presented to measure the importance and popularity of a customer review. We also proposed a customer segmentation approach based on the sentiment analysis and our improved algorithm, the variable-scale clustering (VSC) algorithm. The functionality of the method was verified herein using the iPhone X short-textual data crawling from Weibo, one of the major social media in China, between 18 Jun 2018 and 24 Jun 2018. Customer-centered marketing strategies and customer personalization marketing campaigns were designed through this case study. Moreover, an evaluation measurement of marketing plans was also established to support marketers decide the final candidate plans under funding constraints. Our proposed methods enrich the intention of the current customer relationship management with the help of the popular social media, which could not only develop the scale transformation thought of clustering algorithms in theory, but also save enterprises' time and cost of the marketing decision process in practice.

However, excessive spam messages on the social media provide challenges for our proposed methods. The identity authentication also limits the amount of the available online data. Therefore, future study will focus on the data gathering phase to better fit the data environment of social media. And more contrast experiments will also be implemented on other popular social media (like WeChat and QQ) to further verify the effectiveness of our proposed methods.

Acknowledgments

The research is supported by the national natural science fund of China (71272161).

6 REFERENCES

- [1] Antoine, G. (2018). Uncertainty, risk aversion and international trade. *Journal of International Economics*, 115(1), 145-158. <https://doi.org/10.1016/j.jinteco.2018.09.001>
- [2] Marcus, V. B. (2018). Political risk and the equity trading costs of cross-listed firms. *The Quarterly Review of Economics and Finance*, 69(1), 232-244. <https://doi.org/10.1016/j.qref.2018.03.004>
- [3] Udo, B. & Soumyatanu, M. (2017). International trade and firms' attitude towards risk. *Economic Modelling*, 64(1), 69-73. <https://doi.org/10.1016/j.econmod.2017.03.006>
- [4] Jeong, B., Yoon, J. & Lee, J. (2017). Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. *International Journal of Information Management*. <https://doi.org/10.1016/j.ijinfomgt.2017.09.009>
- [5] Wajda, W. (2019). Innovation, sustainable HRM and customer satisfaction. *International Journal of Hospitality Management*, 76(1), 102-110. <https://doi.org/10.1016/j.ijhm.2018.04.009>
- [6] Radojevic, T., Staniscic, N., Stanic, N. & Davidson, R. (2018). The effects of traveling for business on customer satisfaction with hotel services. *Tourism Management*, 67(1), 326-341. <https://doi.org/10.1016/j.tourman.2018.02.007>
- [7] Zhao, X., Zhan, M. & Liu, B. F. (2018). Disentangling social media influence in crises: Testing a four-factor T model of social media influence with large data. *Public Relations Review*, 44(4), 549-561. <https://doi.org/10.1016/j.pubrev.2018.08.002>

- [8] Shen, C., Chenb, M. & Wang, C. (2018). Analyzing the trend of O2O commerce by bilingual text mining on social media. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2018.09.031>
- [9] Injadat, M., Salo, F. & Nassif, A. B. (2016). Data mining techniques in social media: A survey. *Neurocomputing*, 214(1), 654-670. <https://doi.org/10.1016/j.neucom.2016.06.045>
- [10] Ren, X., Song, M., Haihong, E., & Song, J. (2017). Social media mining and visualization for point-of-interest recommendation. *The Journal of China Universities of Posts and Telecommunications*, 24(1) 67-76. [https://doi.org/10.1016/S1005-8885\(17\)60189-4](https://doi.org/10.1016/S1005-8885(17)60189-4)
- [11] Shen, C. & Kuo, C. (2015). Learning in massive open online courses: Evidence from social media mining. *Computers in Human Behavior*, 51(2), 568-577. <https://doi.org/10.1016/j.chb.2015.02.066>
- [12] Thomaz, G. M., Biz, A. A., Bettoni, E. M., Luiz, M. & Buhalis, D. (2017). Content mining framework in social media: A FIFA world cup 2014 case analysis. *Information & Management*, 54(6), 786-801. <https://doi.org/10.1016/j.im.2016.11.005>
- [13] Zhao, Y., Xu, X. & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence on hotel online textual reviews. *International Journal of Hospitality Management*, 76(1), 111-121. <https://doi.org/10.1016/j.ijhm.2018.03.017>
- [14] Guo, Y., Barnes, S. J. & Jia, Q. (2017). Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59(1), 467-483. <https://doi.org/10.1016/j.tourman.2016.09.009>
- [15] Geetha, M., Singha, P. & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels- An empirical analysis. *Tourism Management*, 61(1), 43-54. <https://doi.org/10.1016/j.tourman.2016.12.022>
- [16] Wei, C. (2012). Investigation and upgrade of customer satisfaction. *Mechanical Management and Development*, 1(1), 185-186.
- [17] Helia, V. N., Abdurrahman, C. P. & Rahmillah, F. I. (2018). Analysis of customer satisfaction in hospital by using Importance-Performance Analysis (IPA) and Customer Satisfaction Index (CSI). *2nd Int. Conf. Eng. Technol. Sustain. Dev*, 154(1), 1-5. <https://doi.org/10.1051/mateconf/201815401098>
- [18] Mohammadian, M. (2017). Intelligent Decision Making and Risk Analysis of B2c E-Commerce Customer Satisfaction. *Fuzzy Syst. Concepts. Methodol. Tools, Appl.*, 2(1), 969-986. <https://doi.org/10.4018/978-1-5225-1908-9.ch042>
- [19] Fasanghari, M. & Roudsari, H. P. F. (2008). The Fuzzy Evaluation of Ecommerce Customer Satisfaction. *World Applied Sciences Journal*, 4(1), 164-168.
- [20] Liu, X., Zeng, X., Xu, Y., & Koehl, L. (2008). A fuzzy model of customer satisfaction index in e-commerce. *Mathematics and Computers in Simulation*, 77(5-6), 512-521. <https://doi.org/10.1016/j.matcom.2007.11.017>
- [21] Yang, S., Zhang, R. & Liu, Z. (2009). The AGA Evaluating Model of Customer Loyalty Based on E-commerce Environment. *Journal of Software*, 4(1), 262-269. <https://doi.org/10.4304/jsw.4.3.262-269>
- [22] Ding S., Wang Z., Wu D., & David L. O. (2017). Utilizing customer satisfaction in ranking prediction for personalized cloud service selection. *Decis. Support Syst*, 93(1), 1-10. <https://doi.org/10.1016/j.dss.2016.09.001>
- [23] Ding S., Yang S., Zhang Y., Liang C., & Xia C. (2014). Combining QoS prediction and customer satisfaction estimation to solve cloud service trustworthiness evaluation problems. *Knowledge-Based Syst.*, 56(1), 216-225. <https://doi.org/10.1016/j.knosys.2013.11.014>
- [24] Qian, X., Li, M., Ren, Y., & Jiang, S. (2018). Social media based event summarization by user-text-image co-clustering. *Knowledge-Based Syst.* <https://doi.org/10.1016/j.knosys.2018.10.028>
- [25] Khalil, S. & Fakir, M. (2017). RCrawler: An R package for parallel web crawling and scraping. *SoftwareX*, 6(1), 98-106. <https://doi.org/10.1016/j.softx.2017.04.004>
- [26] Butakov, N., Petrov, M. & Radice, A. (2016). Multitenant approach to crawling of online social networks. *Procedia Comput. Sci.*, 101(1), 115-124. <https://doi.org/10.1016/j.procs.2016.11.015>
- [27] Howells, K. & Ertugan, A. (2017). Applying fuzzy logic for sentiment analysis of social media network data in marketing. *Procedia Computer Science*, 120(1), 664-670. <https://doi.org/10.1016/j.procs.2017.11.293>
- [28] Krishnamurthy, B., Puri, N., & Goel, R. (2016). Learning Vector-Space Representations of Items for Recommendations using Word Embedding Models. *Procedia Comput. Sci.*, 80(1), 2205-2210. <https://doi.org/10.1016/j.procs.2016.05.380>
- [29] Hiram C., Oscar M., & Marco A. (2016). Integrated concept blending with vector space models. *Comput. Speech Lang.*, 40(1), 79-96. <https://doi.org/10.1016/j.csl.2016.01.004>
- [30] Pham, D. & Le, A. (2018). Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis. *Int. J. Approx. Reason.*, 103(1), 1-10. <https://doi.org/10.1016/j.ijar.2018.08.003>
- [31] Yu, L., Lee, L., Yeh, J., Shih, H., & Lai, Y. (2016). Near-synonym substitution using a discriminative vector space model. *Knowledge-Based Syst.*, 106(1), 74-84. <https://doi.org/10.1016/j.knosys.2016.05.025>
- [32] Stein, R. A., Jaques, P. A., & Valiati, J. F. (2019). An analysis of hierarchical text classification using word embeddings. *Inf. Sci.*, 471(1), 216-232. <https://doi.org/10.1016/j.ins.2018.09.001>
- [33] Mikolov T., Chen K., Corrado G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *Comput. Sci.*, 1(1), 1-12
- [34] Zhang, D., Xu, H., Su, Z., & Xu, Y. (2015). Chinese comments sentiment classification based on word2vec and SVM^{perf}. *Expert Syst. Appl.*, 42(4), 1857-1863. <https://doi.org/10.1016/j.eswa.2014.09.011>
- [35] Khatua, A., Khatua, A., & Cambria, E. (2019). A tale of two epidemics: Contextual Word2Vec for classifying twitter streams during outbreaks. *Inf. Process. Manag.*, 56(1), 247-257. <https://doi.org/10.1016/j.ipm.2018.10.010>
- [36] Carrasco, R. S. M. & Sicilia, M. (2018). Unsupervised intrusion detection through skip-gram models of network behavior. *Comput. Secur.*, 78(1), 187-197. <https://doi.org/10.1016/j.cose.2018.07.003>
- [37] Du, N. (2018). *Study on short text data mining based on social media*. Tianjin: Tianjin University of Technology.
- [38] Zhang, S., Wei, Z., Wang, Y., & Liao, T. (2018). Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary. *Futur. Gener. Comput. Syst.*, 81(1), 395-403. <https://doi.org/10.1016/j.future.2017.09.048>
- [39] Wang, T., He, J., & Wang, X. (2018). An information spreading model based on online social networks. *Physica A: Statistical Mechanics and its Applications*, 490(1), 488-496. <https://doi.org/10.1016/j.physa.2017.08.078>
- [40] Chen, H., Liu, J., Lv, Y., Li, M. H., Liu, M., & Zheng, Q. (2018). Semi-supervised clue fusion for spammer detection in SinaWeibo. *Inf. Fusion.*, 44(1), 22-32. <https://doi.org/10.1016/j.inffus.2017.11.002>
- [41] Gao, X. & Wang, A. (2018). Variable-scale clustering, 2017 *Int. Conf. Logist. INFORMATICS Serv (LISS' 2018)*. In press.
- [42] Zhang, Q., Wang, G., & Hu, J. (2013). *Multi-granularity knowledge acquisition and uncertainty measurement*. Beijing: Science Press, 1(1), 16-25.

Contact information:

Ai WANG, PhD student
(Corresponding author)
University of Science and Technology Beijing
No. 30 Xueyuan Road, Haidian District, Beijing, China
wangai2222@126.com

Xuedong GAO, Professor
University of Science and Technology Beijing
No. 30 Xueyuan Road, Haidian District, Beijing, China
gaoxuedong@manage.ustb.edu.cn