Received: 19 September 2017 | Revised: 23 September 2018 | Accepted: 21 December 2018
DOI: 10.1111/exsy.12374

ORIGINAL ARTICLE

WILEY Expert Systems

Developing a personal value analysis method of social media to support customer segmentation and business model innovation

Tsung-Yi Chen[1] | Hsiang-An Cheng[2] | Yuh-Min Chen[2]

[1] Department of Information Management, Nanhua University, Chiayi, Taiwan (ROC)

[2] Institute of Manufacturing Information and Systems, National Cheng Kung University, Tainan City, Taiwan (ROC)

Correspondence
Tsung-Yi Chen, Department of Information Management, Nanhua University, No. 55, Sec. 1, Nanhua Rd., Dalin Township, Chiayi County 622, Taiwan (ROC).
Email: tsungyi@mail.nhu.edu.tw

Funding information
Ministry of Science and Technology, Grant/Award Numbers: MOST 107-2221-E-343-003-MY3 and MOST 103-2221-E-343-003-MY3

Abstract

Companies need to find out about the personal values of customers and identify customer segments before developing effective business models (BMs) or marketing strategies. Therefore, understanding the personal values of customers is critical in BM design. Traditional personal value evaluation methods are laborious and time-consuming. In recent years, social media have become an important platform for people to share ideas and express views, making the presence of a huge amount of opinions on platforms that can cater for the demand of data for personal value analysis of customers. This study took Facebook and Instagram as the targets and developed a novel personal value forecasting method to help enterprises obtain the various personal values of customer segments automatically at a lower cost. This study adopted Schwartz's value theory as the value model and proposed a consistency model and a relativity model for weighted calculations, so as to determine the feature of a value tag. Finally, the feature selection algorithm and classification algorithm were used for judging values. In the evaluation phase of this study, 61 participants were recruited to test the proposed method. The proposed method could assist enterprises in better understanding personal value information.

KEYWORDS

association rule, business model, personal values, Schwartz value survey, social media

1 | INTRODUCTION

The primary causes for business failure are failing to meet market demand, lacking an appropriate business model (BM), and neglecting customer segments. In the course of market competition, many factors are involved to enhance competitiveness. Aside from products, technology, service, brand, and corporate culture, BM innovation is also important. In other words, BM innovation is the actual driving force that determines an enterprise's innovation value. In a highly competitive environment, lacking specific information about consumer psychology may result in business failure. Therefore, in order to design a complete and effective BM, the features of customer segments must be known first.

Enterprises should utilize the information of their customers to understand and consider the specific preferences and psychological traits of their customers. Therefore, customers' personal value is an important feature that should be understood. Personal values represent the acceptable activities, beliefs, products, and services of a person. Personal values not only influence individuals' tastes and preferences for specific products but also impact the final decision. Therefore, the information of personal values helps enterprises identify customer segments.

Past literature has discussed personal value theory and conducted expert interviews or used a personal value scale to collect data. Although it is effective when the sample pool is small, for enterprises that face numerous customer segments, scale surveys or interviews are not feasible. Therefore, when designing BMs, an automated method or tool for analysing the personal values of numerous customer segments is necessary.

There are presently two types of research on personal values: (a) using a personal value scale to survey the subjects' values, so as to know the personal values in different countries (Sagiv & Schwartz, 2000; Schwartz, 1992; Swartz, 1996) and (2) improving the personal value scale according to individual demand by analysing and verifying the results (Krystallis, Vassallo, & Chryssohoidis, 2012). In recent years, social media (Stieglitz, Mirbabaie, Ross, & Neuberger, 2018; Tajudeen, Jaafar, & Ainin, 2018; Wikipedia, 2016a) such as Facebook and Instagram have become platforms for users to offer their personal opinions, which can be viewed by the general public (Golbeck, Robles, & Turner, 2011). Gou, Zhou, and Yang (2014) mentioned that there is research on using the digital footprints of users on social media to predict personal traits and gain a deeper understanding of users. Some scholars (Kosinski, Stillwell, & Graepel, 2013Ortigosa, Martín, & Carro, 2014Youyou, Kosinski, & Stillwell, 2015) have used Facebook data to predict users' recessive traits, but most of them employed questionnaires to determine the types of users before the analysis based on user data. Few studies have discussed information of recessive traits before utilizing the obtained information for analysis.

This study proposed a novel personal value forecasting method to help enterprises obtain the various personal values of customer segments automatically at a lower cost, so as to assist them in designing BMs. This study used Facebook as the experimental target and took the Chinese processing status as the research objective. In order to extensively obtain personal value tags for the experiment, this study also collected data from Instagram. This study then inferred the personal value type from the statuses selected by the Facebook users.

As it was difficult to find experimental subjects, there were only 61 participants to test the proposed method. The feature calculated by the proposed method was used to employ the feature selection algorithm and classification algorithm (Hall et al., 2009) and utilized tenfold cross-validation to obtain accuracy. The proposed method could assist enterprises to better understand personal value information, in order to determine customer segments and design a more accurate and more effective BM.

2 | RELATED RESEARCH

2.1 | Business models, social media, and association rule mining

BM is essential to enterprises, and many enterprises rely on BM for successful operation. Fritscher and Pigneur (2009) proposed the Business Model Canvas (BMC) that is a method that systematically combines BMs for describing how an organization creates, transfers, and obtains value. BMC consists of nine elements: customer segments, value propositions, channels, customer relationships, revenue streams, key resources, key activities, key partnerships, and cost structure. BMC is one of the most extensively used methods in commercial applications (Flodén & Williamsson, 2016; Kaplan, 2012) or theory-oriented academic research (Ojasalo & Ojasalo, 2015Toro-Jarrín, Ponce-Jaramillo, & Güemes-Castorena, 2016).

Social media are forms of electronic interaction. Through social media, users can build online communities to share ideas, opinions, information, messages, and other content types in different formats (Jabeur, Nait-Sidi-Moh, & Zeadally, 2017). The rise of social media usage has opened up new opportunities for analysing data to gain patterns, insights, or trends (Stieglitz et al., 2018). Gou et al. (2014) discovered the potential feasibility and effectiveness of automatically deriving user personality traits from Twitter. Chen, Hsieh, Mahmud, and Nichols (2014) analysed people's values and word use on an online social news sharing community (Reddit). Social media help organizations and businesses to understand customer needs and motivates them to respond (Tajudeen et al., 2018).

Current research on analysing social media users includes analysing user interest from social media (Bhargava, Brdiczka, & Roberts, 2015), using analytic social media to understand user personality traits (Chen et al., 2014; Skowron, Tkalčič, Ferwerda, & Schedl, 2016Wei et al., 2017), using topic tags to predict user brand preference prediction of community media (Yang, Pan, Mahmud, Yang, & Srinivasan, 2015), or using social media data to predict consumption behaviour (Zhang & Pennacchiotti, 2013).

Mark Zuckerberg established Facebook as a network social media platform in 2004, and since then, the number of active users has increased to about 1.6 billion (Linli, 2016). One of the core functions of Facebook is Timeline, where a user can post a message or a video on his or another's Facebook page. Any Facebook user can respond to a post, and a post issued by the user in Timeline is called a Status. Another core function is Like, where the liked object can be one's own post or another person's post; even firms, stores, public figures, and concepts can set up a Page. Other functions, such as Check-in, creating albums, writing notes, transferring messages, and creating events, enable users to break space-time limitations and establish contact with friends in the network more conveniently. Instagram, another social media platform, is characterized by sharing pictures and videos. The user can issue a post on a personal page, and each post can be liked and commented on. The user can track others to obtain the latest information, which is a simple function as compared with Facebook.

A hashtag is a practical function that is represented as a # symbol in front of a single word or a sentence without spaces. Hashtags allow posts of related subjects or events to be gathered together (Small, 2011). For example, when the hashtag "#pleasure" is clicked, those posts with a "#pleasure" tag will be displayed. Hashtags are also applicable to both Facebook and Instagram, in which a post obtained through a hashtag is known as a hashtag post. The content in the hashtag posts of Facebook is more abundant than that in the hashtag posts of Instagram. The hashtag posts of Facebook contain articles, likes, shares, hashtags, comments, and media. The hashtag posts of Instagram contain articles, likes, hashtags, comments, and media. In the environment of social media, hashtags can be used to find a specific subject or content and can be used for the dissemination and searching of messages, event detection, and subject analysis (Zhao, Zhu, Jin, & Yang, 2015).

Association rule mining is a useful method for discovering interesting relations and dependencies between variables in large databases (Feng, Cho, Pedrycz, Fujita, & Herawan, 2016; Sawant & Shah, 2016). Three algorithms, such as Apriori, AprioriTid, and AprioriHybrid, are frequently used to extract association rules from databases (Feng et al., 2016). Some recent studies have analysed hashtags. For example, Adedoyin-Olowe, Gaber, Dancausa, Stahl, and Gomes (2016) used hashtags and association rules to explore valuable news. Hamed, Wu, Erickson, and Fandy (2015) examined hashtags and association rules to find the correlation among drugs. ElTayeby, Molnar, and George (2014) investigated hashtags in social media and used association rules to discuss opinions in the network.

2.2 | Schwartz value theory

Personal values are the decision criteria people use when judging between right and wrong when making selections (Wikipedia, 2016b). They are a long-term and stable mental state that are unlikely to be changed by any external influence (Schein, 1985). Personal values have been shown to influence a wide range of human behaviours (Chen et al., 2014). The evaluation results of personal values can be used for long-term forecasts and are important for market decision making (Krystallis et al., 2012). Therefore, this study analysed the types of personal value according to the explicit behaviour derived from people's values. Explicit behaviour was defined as the status published by the user in Facebook.

Many scholars have proposed theories of classification for personal values, with common personal value theories including the Rokeach value survey (Rokeach, 1973), Hofstede value survey (Hofstede, 1980), and Schwartz value survey (Schwartz, 1992). The advantages of the Schwartz value survey in comparison to the other two theories are the following: (a) the personal value questionnaire is designed for multiple cultures and covers 60 countries and regions, including Taiwan (Schwandt, 1994Schwartz, 1992Schwartz & Sagiv, 1995); and (b) the personal value scale is merged with the Rokeach value questionnaire, the Chinese value scale, and the Hofstede value scale; it is a mature value scale that has evolved from various cultures. Therefore, this study used the Schwartz value survey as the basis of value classification.

The structure and circle diagrams of the Schwartz value survey are shown in Figures 1 and 2 respectively. The Schwartz value survey is used to reduce specific value items (SVIs) into 10 classes of personal values, and the values are reduced to value dimensions (high-order value) that are dispersed over two orthogonal axes. In the circle diagram, the horizontal axis represents people's conservation (CO) and openness to change (OP). The vertical axis represents human's self-transcendence (ST) and self-enhancement (SE). The pairwise high-order value can generate four quadrant values that are self-transcendence-openness to change (STOP), self-transcendence-conservation (STCO), self-enhancement-conservation (SECO), and self-enhancement-openness to change (SEOP), meaning individuals' values may fall into one of the four quadrant values. The four high-order personal values are (a) self-transcendence (ST), which is characterized by tolerance, acknowledgement, allowance, protecting natural resources, enhancing human welfare, and protecting collective welfare; (b) self-enhancement (SE), which is characterized by social class and fame, controlling various relations and power, and procuring personal achievement without breaking social norm; (c) openness to change (OP), which is characterized by independent thinking and autonomous selection, creation, and discovery activities (self-direction), and being satisfied with experiencing different stimulations and challenges in life; and (d) conservation (CO), which is characterized by restraining one's own behaviour from damaging others' interests or social norms, ideas, or impulses, respecting cultural traditions or religious customs, stable personal, interpersonal, and social

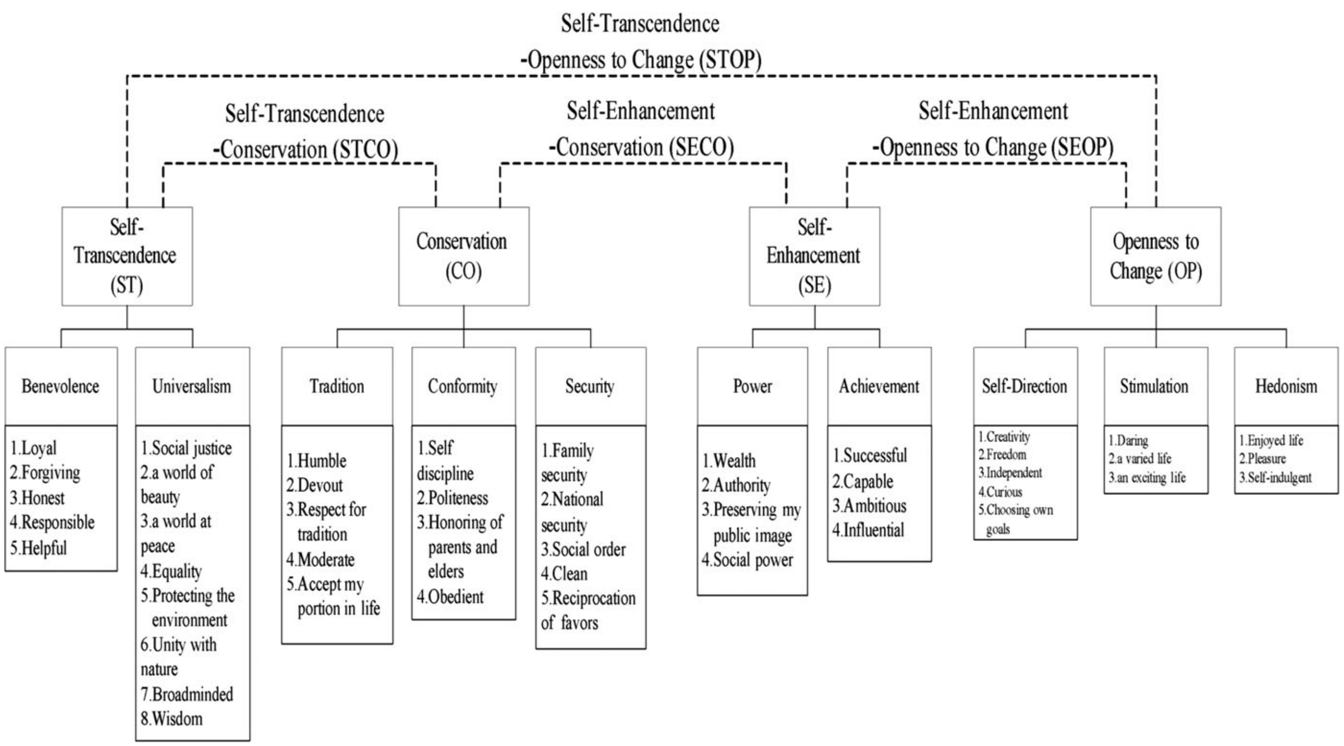

FIGURE 1 Structure of Schwartz value theory

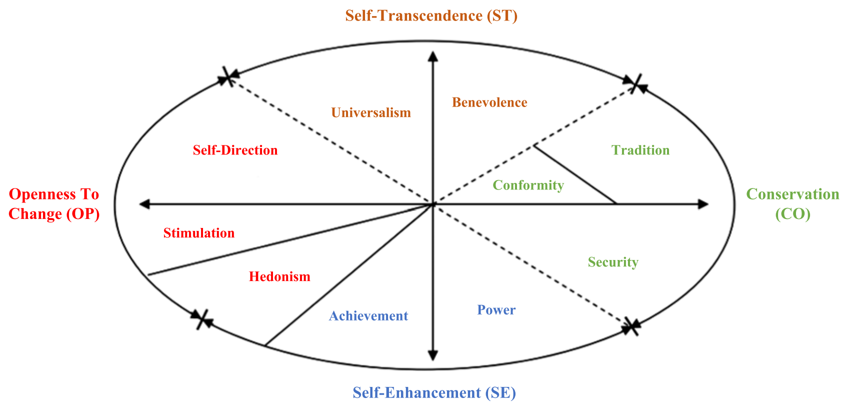

FIGURE 2 Circle diagram of Schwartz value theory

relations, easement, and harmony. This study used the Portrait Values Questionnaire (PVQ) proposed by Schwartz (1992Swartz, 1996; Sagiv & Schwartz, 2000; Krystallis et al., 2012) to evaluate the subjects' personal value type.

3 | METHODS

Users with different personal values have different ways to express self-particularity to the outside. The method proposed in this study was expected to deduce the personal value type from the prototype implied in the Chinese statuses of Facebook users. First, a four-stage personal value analysis method was designed (Figure 3) below.

1. Generating a personal value tag: SVI is used as the tag, the hashtag post correlated with the tag is extracted, and the value tag of each type of value is deduced from the tag contained in the hashtag post by using the improved association rules proposed herein.
2. Building the term weight table of the value tags: A post with a value tag is extracted from the value tag obtained in the previous stage as the feature article of the value tag. Afterwards, the augmented normalized term frequency (Salton & Buckley, 1988) and the consistency model and relativity model designed by this study are used to weight the term, so as to generate the term weight table of the value tag.
3. Building the term weight table of the user: The status of the Facebook user is obtained, and then the term weight table of the user is built, so as to obtain the feature of the user.
4. Value type analysis of the user: The weight table of the value tag obtained in Stage 2 is used to analyse the user weight table obtained in Stage 3, and each user's score for each value tag is calculated. Finally, the feature selection algorithm and classification algorithm (Hall et al., 2009) are used for analysis. The feature selection algorithm is the process of selecting a subset of relevant features that minimizes the error rate (Guyon & Elisseeff, 2003).

3.1 | Generating personal value tags

In order to analyse the particularity of Facebook users' status, the particularity of each type of personal value must be found as the feature. Sagiv and Schwartz (2000) proposed the SVI for each type of personal value. SVI refers to the particularity involved in the corresponding personal values. The method in this stage is detailed in Sections 3.1.1 and 3.1.2.

3.1.1 | Hashtag post capture

As this study conducted the experiment in Chinese, the problems concerning Chinese SVI were discussed (Tian, 2008; Wang, 2012). There were two problems when the present SVI proposed by scholars was used in this study: (a) Chinese SVI names are not unified; and (b) not every SVI can find numerous hashtag posts. Therefore, according to (a), the Chinese SVI used in this study was generated without influencing the original intention of the SVI, and according to (b), the SVI of over 500 hashtag posts was included in this study.

The tags of the hashtag posts were then extracted based on SVI. For example, the hashtag post "#blood donation #enthusiasm #commonwealth #charitable deeds #physical health haven't donated blood for long, donate some blood for commonwealth" was obtained by using the Chinese SVI "#commonwealth," extracting the tags "#blood donation #enthusiasm #commonwealth #charitable deeds #physical health" of the hashtag post and then extracting the tags of multiple hashtag posts to generate the itemsets of the SVI.

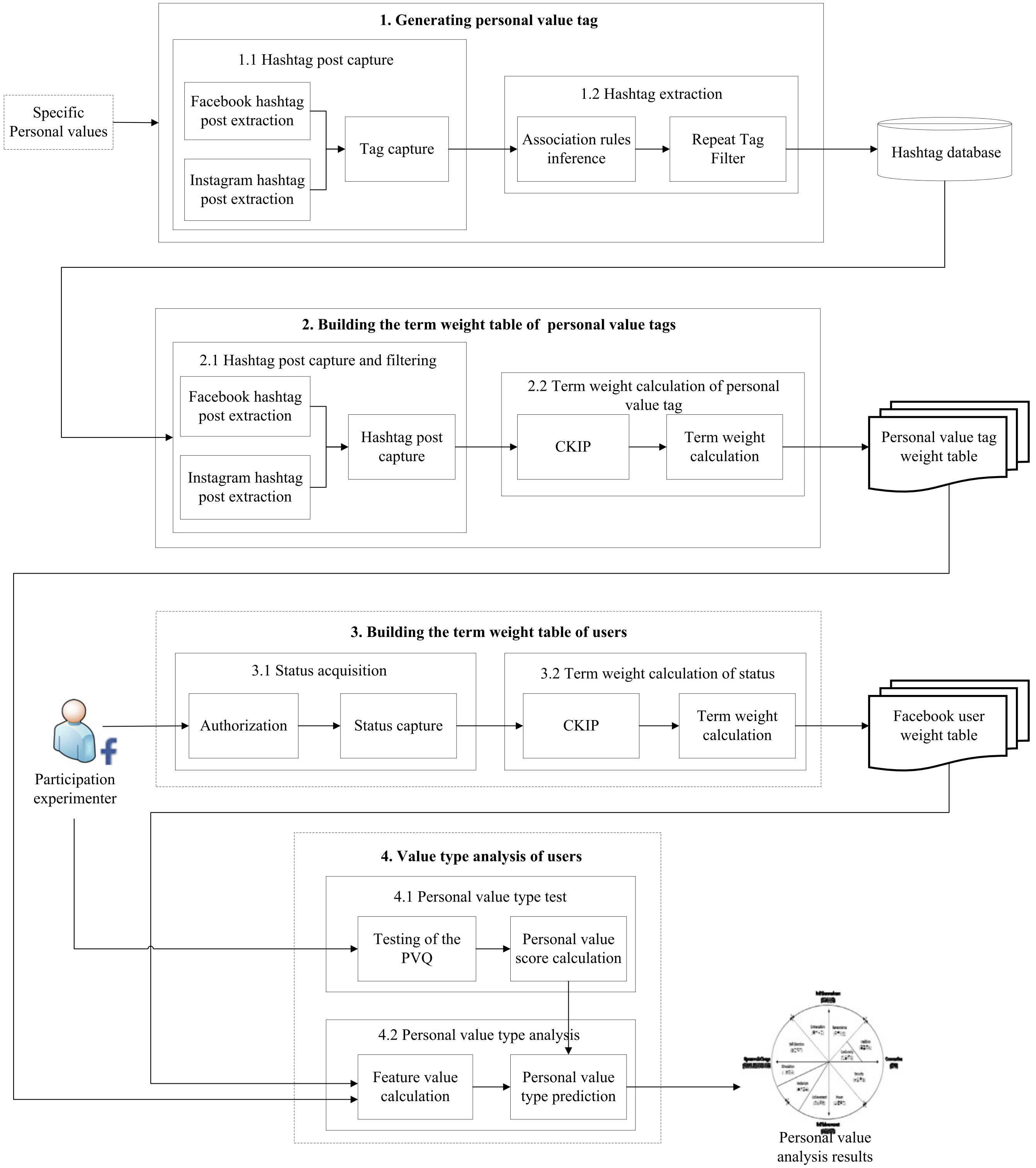

FIGURE 3 Personal value analysis method

3.1.2 | Personal value tag extraction

The itemsets of the SVI obtained in the previous step were deduced using the association rule. The Apriori algorithm proposed by Agrawal and Srikant (1994) is the most typical algorithm of the association rule. However, there were two defects in using the Apriori algorithm in this study, which reduced the computational efficiency: (a) when the Apriori algorithm generates candidate 2-itemsets from the frequent 1-itemsets, numerous unwanted candidate itemsets will be generated; and (b) as this study searched the personal value tags of SVI, the frequent 2-itemsets

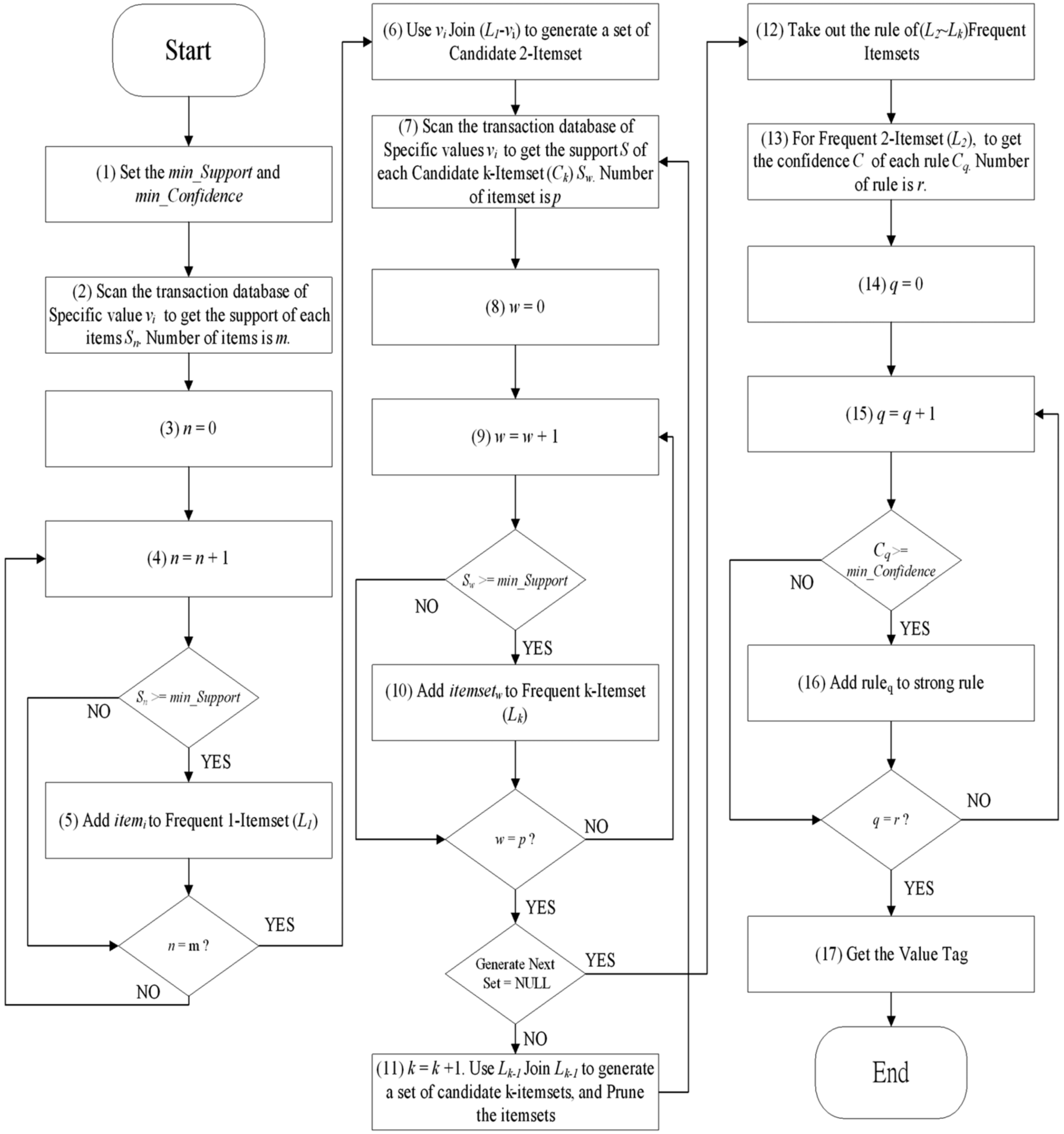

FIGURE 4 Improved association rule algorithm

were predominant and other frequent itemsets in different lengths were used to judge the weight in the case of an identical tag. Therefore, an improved association rule algorithm was designed first (Figure 4). The procedure is described below.

(S1). Set the *min_Support* and *min_Confidence* for the association rule.

(S2). Scan the itemsets generated by *SVI* (v_i). Set the number of items as m and calculate the Support S_n of each item. The computing mode is the occurrence number of items in the database, $n = 1 \sim m$.

(S3). Initialize n.

(S4). Add 1 to n, and then judge if S_n is greater than or equal to the preset *min_Support*; proceed to (S5) if yes; otherwise, judge if n is equal to m, that is, whether it is the last item; proceed to (S6) if yes; proceed to (S4) if no; continue to calculate the next item.

(S5). When the support of item S_n is greater than or equal to the preset *min_ Support*, the item $item_n$ is added in the frequent 1-itemset (L_1) in length k of 1. Afterwards, judge if n is equal to m; proceed to (S6) if yes; proceed to (S4) if no, and continue to calculate the next item.

(S6). Use frequent 1-itemset (L_1) in length k of 1 to generate candidate 2-itemset (C_2) in length k of 2 and then proceed to (S7).

(S7). Scan the itemsets generated from *SVI* (v_i). The number of items is p, and the support S_w of candidate k-itemset (C_k) in length k is calculated according to the occurrence number of itemsets in database; $w = 1 \sim p$.

(S8). Set w as 0.

(S9). Add 1 to w and then judge if S_w is greater than or equal to the preset *min_Support*, proceed to (S10) if yes; otherwise, judge if w is equal to p and proceed to (S9) if no to calculate the next item. If yes, judge if new itemsets can be generated; proceed to (S11) if no, or proceed to (S12) if yes.

(S10). When the support of itemset S_n is greater than or equal to the set *Min_ Support*, the itemset $item_w$ is added in the frequent k-itemset (L_k) in length k. Afterwards, judge if w is equal to p or not, and proceed to (S9) to calculate the next item if no; judge whether new itemsets can be generated if yes; proceed to (S11) if no, or proceed to (S12) if yes.

(S11). Add 1 to k. Unite frequent itemset L_{k-1} and L_{k-1} to generate candidate k-itemsets and use the Prune algorithm to delete the discrepant items. Return to (S7).

(S12). Extract each rule in length meeting *Min_Support*.

(S13). Calculate the reliability of frequent 2-itemset (L_2) rule and scan the itemsets database generated from *SVI* (v_i). The number of rules is p, which meets the length generated by confidence C_q of the rule. If the deduced rule is X- > Y, then the reliability represents the proportion of X in the data $X \cup Y$, $q = 1 \sim r$.

(S14). Set q as 0.

(S15). Add 1 to q, and judge if C_q is greater than or equal to *min_Confidence*; proceed to (S16) if yes; judge whether q is equal to r; if no, proceed to (S15) to calculate the next rule; proceed to (S16) if yes.

(S16). Add $rule_q$ in the strong rule.

(S17). Extract the tags from the strong rule, known as the personal value tags.

The same personal value tag may be deduced from different SVIs, and so a method for judging the personal value tag attribution is designed (Equation 1):

$$MAX(V(i_k,j)) = MAX\left(\frac{\sum_{l=2}^{N} S(i_k,j,l)^* \log_2 l}{I_k}\right), \quad k = 1\tilde{\ }m, \tag{1}$$

where $V(i_k,j)$ is the effective personal value of Tag j deduced from the i_k SVI, and $S(i_k,j,l)$ is the support value of value tag j in frequent itemsets in length l deduced from the i_k SVI. The support is multiplied by the *log* value of l, $k = 1 \sim m$, m is the SVI number of all the deduced value tag j, $log_2 l$ is the purpose of adjusting the weight, and I_k is the number of items contained in value tag k, for calculating the proportion of the support of the item. The maximum value is calculated, and the tag is attributed to SVI.

3.2 | Building the term weight table of personal value tags

The status could not be analysed only with the obtained tags; therefore, more information about personal value tags was required found, and the term weight table of the personal value tag needed to be calculated. The details are described below.

3.2.1 | Hashtag post capture and filtering

First, hashtag posts with a personal value tag were extracted from Facebook and Instagram. This study designed three rules to filter hashtag posts.

1. Non-Chinese-based hashtag posts: If non-Chinese articles were added in, there would be errors in the result.
2. Tag-based hashtag posts: Some hashtag posts may use tags only to express moods, and this type of article might have very little content.
3. Link only hashtag posts: Some hashtag posts only share links, and this type of article would be unhelpful to the result.

Afterwards, the effective posts with the same personal value tag were combined into a personal value tag article (VTA) of the value tag.

3.2.2 | Term weight calculation of the personal value tag

This step found the corresponding keyword group of each personal value tag based on the term weight table calculation of the tag, so as to automatically judge the feature of unknown articles without a tag and to analyse the weight of the article in each personal value tag.

First, CKIP (Chen & Ma, 2005) was used for the word segmentation, sentence segmentation, and part-of-speech tagging of VTA, and the punctuation marks that could influence the result were deleted, so as to increase the computational efficiency. As articles on Facebook and Instagram have different lengths, in order to avoid calculation errors, this study used the augmented normalized term frequency for calculation. The augmented normalized term frequency uses the maximum TF value in the article for normalization, which can be expressed as Equations 2 and 3:

$$aTF_{ij} = 0.5 + 0.5 \times \frac{TF_{ij}}{\mathrm{Max}(TF_j)},\ TF_{ij} = \frac{n_{ij}}{|d_i|}, \tag{2}$$

$$IDF_j = log\frac{N}{n_j}, \tag{3}$$

where TF_{ij} is the term frequency (TF) of term w_j in document d_i; n_{ij} is the occurrence number of term w_j in document d_i; $|d_i|$is the number of terms in document d_i; IDF_j uses associated inverse document frequency (IDF) to represent the importance of w_j in all articles; N is the total number of documents in dataset; and n_j is the number of relevant documents with term w_j.

The computing mode consists of the aTF_{ij} value multiplied by the IDF_j value to obtain the *aTFIDF* value. *Max* (TF_j) is the maximum TF value in the article. The TF_{ij} of the term in the current document is divided by the maximum TF value in the article for normalization, so as to not influence the objective result.

The *aTFIDF* value calculated by the aforesaid method does not consider the terms tagged in the article. The Schwartz value theory consists of two axes: self-transcendence and self-enhancement and openness to change and conservation, which are pairwise opposite. During the calculation of the term weight table of the self-transcendence value tag, if there is an article tagged as the personal value tag of self-enhancement, then the weight of the article is reduced. On the contrary, during the calculation of the term weight table of the self-transcendence personal value tag, if an article is tagged as the personal value tag of self-transcendence, then the weight personal value of the article is increased.

In answer to the aforesaid conditions, the n_{ij} computing mode in TF was improved in this study. For each tagged article, the term weight was enhanced or reduced. The following two models were proposed in this stage.

1. Consistency model: If the term of the article often has a personal value tag identical with this value, then the reliability of the article in the values can be increased. For example, to calculate the term weight table of the “#pleasure” tag, there are “#happy #free” tags of the openness to change personal value dimension such as “#pleasure”; therefore, the weight of the article is enhanced, which can expressed as Equation 4:

$$CTijk = CntOfConTag(wj, ti, s_k, PT),\ s_k \in S_i, \tag{4}$$

where *CntOfConTag* (w_j, t_i, s_k, *PT*) is the number of tags of term w_j tagged as the value dimension, which is the same as tag t_i in s_k articles in article s_i of personal value tag t_i; w_j is the term to calculate the weight; t_i is the personal value tag; s_i is the VTA of term t_i; s_k is the subarticle in article s_i; and *PT* is a positive term, representing the term set of the same personal value dimension as term t_i.

2. Relativity model: If the term of the article often has a personal value tag opposite to this value, then the reliability of the article regarding the personal value is reduced. For example, to calculate the term weight table of the “#pleasure” term, there are “#humbleness #devoutness” terms of personal value dimension opposite to “#pleasure”; therefore, the weight of the article is reduced, which can be expressed as Equation 5:

$$REijk = CntOfRETag(wj, ti, s_k, NT),\ s_k \in S_i, \tag{5}$$

where *CntOfRETag* (W_j, t_i, s_k, *NT*) is the term frequency of term w_j tagged as the personal value dimension identical with term t_i in s_k articles in article s_i of personal value tag t_i; w_j is the term to calculate the weight; t_i is the personal value tag; s_i is the VTA of term t_i; s_k is the subarticle in article s_i; and *NT* is a negative term, representing the term set of the personal value dimension opposite to term t_i.

According to the aforesaid two models, the computing mode for n_{ij} of the TF value is expressed as Equations 6 and 7:

$$OCnt_{jk} = CntofOccur(wj, s_k), \tag{6}$$

$$n_{ij} = \sum_{k=1}^{m} PNWt_{ijk}, \tag{7}$$

$$PNWt_{ijk} = \begin{cases} OCnt_{jk} & \text{if } CT_{ijk} - RE_{ijk} - 1 = 0 \\ OCnt_{jk} * \log_2\left(\left(CT_{ijk} + RE_{ijk}\right) + 1\right) & \text{if } CT_{ijk} - RE_{ijk} - 1 > 0 \\ OCnt_{jk} / \log_2\left(\left(CT_{ijk} + RE_{ijk}\right) + 2\right) & \text{if } CT_{ijk} - RE_{ijk} - 1 < 0 \end{cases},$$

where *CntofOccur* (w_j, s_k) is the occurrence number of term w_j in article s_k; n_{ij} is the occurrence number of term w_j in the VTA generated by personal value tag t_i in the original computing mode, and the characteristic of the personal value is added after improvement for weighting; m is the number of documents contained in VTA, that is, $|S_i|$; and PNW_{ijk} is the occurrence number of term w_j in the VTA generated by personal value tag t_i after weighting.

The weighted *TF* is integrated and combined with the computing equation for term weight after *IDF* (Equation 8):

$$Weight_{vij} = \left(0.5 + 0.5 * \frac{\sum_{k=1}^{m} PNWtijk}{|di| * MAX(TF_j)}\right) * log\frac{N}{nj}, \tag{8}$$

where $Weight_{vij}$ is the weight of term w_j of the ith tag in the v personal value dimension.

After the above calculation, each personal value dimension can generate a weight array of tags and terms (Figure 5). Each element $Weight_{vij}$ in the array is the weight of the left corresponding term $Term_{v,i}$ corresponding to the upper personal value tag $Tag_{v,j}$. For example, $Weight_{1,2,2}$ is the weight of $Term_{1,2}$ corresponding to $Tag_{1,2}$.

3.3 | Building the term weight table of users

The most direct way for the user to show himself in Facebook is to publish his status in the Timeline. Thus, the user's status is obtained first at this stage, and then the corresponding term weight table is generated. Below are the details.

3.3.1 | Status acquisition

Facebook provides an application programming interface (API) for the user to obtain related data. Therefore, API can be used to obtain a user's status. The user's authorization must be obtained first (called the Access Token), so as to obtain the data in Facebook. Afterwards, the user's status in the Timeline is extracted.

3.3.2 | Term weight calculation of status

The statuses of the same user are gathered together and called the user articles, and then the TF of TFIDF is used for weight calculation (Equation 9):

$$TF_{ij} = \frac{n_{ij}}{|d_i|}, \tag{9}$$

where TF_{ij} is the TF of term w_j in document d_i, n_{ij} is the occurrence number of term w_j in document d_i,$|d_i|$is the number of terms of document d_i, and d_i is the user article.

After the aforesaid calculation, a weight array of users and terms can be generated (Figure 6), and each element TF_{ij} in the array is the weight of the upper corresponding term $Term_i$ corresponding to the left user $User_j$. For example, $TF_{2,\,2}$ is the weight of $Term_2$ corresponding to $User_2$.

$$\begin{array}{cc} & \begin{array}{cccc} Tag_{1,1} & Tag_{1,2} & \dots & Tag_{1,n} \end{array} \\ \begin{array}{c} Term_{1,1} \\ Term_{1,2} \\ \vdots \\ Term_{1,m} \end{array} & \begin{bmatrix} Weight_{1,1,1} & Weight_{1,1,2} & \dots & Weight_{1,1,n} \\ Weight_{1,2,1} & Weight_{1,2,2} & \dots & Weight_{1,2,n} \\ \vdots & \vdots & \ddots & \vdots \\ Weight_{1,m,1} & Weight_{1,m,2} & \dots & Weight_{1,m,n} \end{bmatrix} \end{array}$$

FIGURE 5 The weight array of personal value tag and term

$$\begin{array}{c c} & \begin{array}{cccc} Term_1 & Term_2 & \cdots & Term_m \end{array} \\ \begin{array}{c} User_1 \\ User_2 \\ \vdots \\ User_u \end{array} & \begin{bmatrix} TF_{1,1} & TF_{1,2} & \cdots & TF_{1,m} \\ TF_{2,1} & TF_{2,2} & \cdots & TF_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ TF_{u,1} & TF_{u,2} & \cdots & TF_{u,m} \end{bmatrix} \end{array}$$

FIGURE 6 The weight array of user and term

3.4 | User's personal value type analysis

The feature data of personal values and users could be obtained through the aforesaid steps. The personal values of users were analysed in this step, and the real personal values of users were tested.

3.4.1 | Personal value type test

This study used the personal value scale of PVQ, with 40 questions proposed by Swartz (1996), in order to increase the validity and reliability of the measurement. In order to reduce the differences among different respondents in reading the scale questions, each question in the scale was provided with a brief context description, so as to guide the subjects to answer independently. Measurement was based on a Likert 6-point scale. The subjects responded according to their personal motives, intentions, and expected similarities.

3.4.2 | Personal value type analysis

The degree of correlation between status and the compared VTA was judged by calculating the similarity score. The similarity score computing mode could be expressed as Equation 10:

$$Simscore_{q,d} = \sum_{i=1}^{n} \left(TF_q(term_i) \times Weight_d(term_i) \right), \tag{10}$$

where q is the status; d is the VTA; $Simscore_{q,d}$ is the similarity score between status q and VTA d, in which a larger value represents higher similarity; T_q is the exclusive term set in status q; T_d is the exclusive term set in VTA d (from the term weight table); $term_i \in T_q \cup T_d$; $n = |T_q \cup T_d|$; $TF_q(term_i)$ is the TF value of $term_i$ in status q; and $Weight_d(term_i)$ denotes the weight value of $term_i$ in the VTA d term set. Each $TF_q(term_i) \times Weight_d(term_i)$ value denotes the contribution score of $term_i$ to the similarity between two articles (q, d). The total score can be the basis of similarity between articles q and d.

The user's feature article calculated the similarity score of all VTAs in this computing mode, and each user could obtain the score of the corresponding personal value tag according to self-transcendence and self-enhancement, as well as openness to change and conservation. As shown in Figure 7, $User_i$ is a Facebook user; i = 1~m, where m is the number of users; $Tag_{j,k}$ is the personal value tag; j = 1~10 representing the total number of personal value dimensions, which are universalism, benevolence, achievement, power, self-direction, stimulation, hedonism, tradition, conformity, and security; k is the number of personal value tags of the personal value dimension; and $Sim_{i,j,k}$ is the similarity score of a user's status and personal value tag. For example, $Sim_{2,1,1}$ is the similarity score of $User_2$ corresponding to personal value tag $Tag_{1,1}$.

For the dimension-based personal value tag and user array in Figure 7, based on the tag, each score is equally divided into high, medium, and low groups, so as to reduce the probability that an excessive score gap may cause a deviation. Figures 8–10 show the array for calculating self-transcendence and self-enhancement, the array for calculating openness to change and conservation, and the array for calculating quadrant values. $User_i$ is a Facebook user; i = 1~m, where m is the number of users; $Tag_{j,k}$ is the personal value tag; j = 1~10, representing the total number of personal value dimensions; k is the number of personal value tags of the personal value dimension; and $Rank_{i,j,k}$ is the level of the Simscore in tag j (high, medium, or low). For example, $Rank_{2,1,1}$ is the rank of user $User_2$ corresponding to personal value tag $Tag_{1,1}$.

$$\begin{array}{c c} & \begin{array}{ccccccccccc} Tag_{1,1} & \cdots & Tag_{1,t1} & Tag_{2,1} & \cdots & Tag_{2,t2} & \cdots & Tag_{10,1} & \cdots & Tag_{10,t10} \end{array} \\ \begin{array}{c} User_1 \\ User_2 \\ \vdots \\ User_u \end{array} & \begin{bmatrix} Sim_{1,1,1} & \cdots & Sim_{1,1,t1} & Sim_{1,2,1} & \cdots & Sim_{1,2,t2} & & Sim_{1,10,1} & \cdots & Sim_{1,10,t4} \\ Sim_{2,1,1} & \cdots & Sim_{2,1,t1} & Sim_{2,2,1} & \cdots & Sim_{2,2,t2} & \cdots & Sim_{2,10,1} & \cdots & Sim_{2,10,t4} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\ Sim_{u,1,1} & \cdots & Sim_{u,1,t1} & Sim_{u,2,1} & \cdots & Sim_{u,2,t2} & & Sim_{u,10,1} & \cdots & Sim_{u,10,t4} \end{bmatrix} \end{array}$$

FIGURE 7 Personal value tag Simscore array of users

$$
\begin{array}{c} \\ User_1 \\ User_2 \\ \vdots \\ User_u \end{array}
\begin{array}{c}
\begin{array}{ccccccccccc} Tag_{1,1} & \cdots & Tag_{1,t1} & Tag_{2,1} & \cdots & Tag_{2,t2} & \cdots & Tag_{4,1} & \cdots & Tag_{4,t4} \end{array} \\
\begin{bmatrix}
Rank_{1,1,1} & \cdots & Rank_{1,1,t1} & Rank_{1,2,1} & \cdots & Rank_{1,2,t2} & & Rank_{1,4,1} & \cdots & Rank_{1,4,t4} \\
Rank_{2,1,1} & \cdots & Rank_{2,1,t1} & Rank_{2,2,1} & \cdots & Rank_{2,2,t2} & \cdots & Rank_{2,4,1} & \cdots & Rank_{2,4,t4} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\
Rank_{u,1,1} & \cdots & Rank_{u,1,t1} & Rank_{u,2,1} & \cdots & Rank_{u,2,t2} & & Rank_{u,4,1} & \cdots & Rank_{u,4,t4}
\end{bmatrix}
\end{array}
$$

FIGURE 8 The array for calculating self-transcendence and self-enhancement

$$
\begin{array}{c} \\ User_1 \\ User_2 \\ \vdots \\ User_u \end{array}
\begin{array}{c}
\begin{array}{ccccccccccc} Tag_{5,1} & \cdots & Tag_{5,t5} & Tag_{6,1} & \cdots & Tag_{6,t6} & \cdots & Tag_{10,1} & \cdots & Tag_{10,t10} \end{array} \\
\begin{bmatrix}
Rank_{1,5,1} & \cdots & Rank_{1,5,t5} & Rank_{1,6,1} & \cdots & Rank_{1,6,t6} & & Rank_{1,10,1} & \cdots & Rank_{1,10,t10} \\
Rank_{2,5,1} & \cdots & Rank_{2,5,t5} & Rank_{2,6,1} & \cdots & Rank_{2,6,t6} & \cdots & Rank_{2,10,1} & \cdots & Rank_{2,10,t10} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\
Rank_{u,5,1} & \cdots & Rank_{u,5,t5} & Rank_{u,6,1} & \cdots & Rank_{u,6,t6} & & Rank_{u,10,1} & \cdots & Rank_{u,10,t10}
\end{bmatrix}
\end{array}
$$

FIGURE 9 The array for calculating openness-to-change and conservation

$$
\begin{array}{c} \\ User_1 \\ User_2 \\ \vdots \\ User_u \end{array}
\begin{array}{c}
\begin{array}{ccccccccccc} Tag_{1,1} & \cdots & Tag_{1,t1} & Tag_{2,1} & \cdots & Tag_{2,t2} & \cdots & Tag_{10,1} & \cdots & Tag_{10,t10} \end{array} \\
\begin{bmatrix}
Rank_{1,1,1} & \cdots & Rank_{1,1,t1} & Rank_{1,2,1} & \cdots & Rank_{1,2,t2} & & Rank_{1,10,1} & \cdots & Rank_{1,10,t10} \\
Rank_{2,1,1} & \cdots & Rank_{2,1,t1} & Rank_{2,2,1} & \cdots & Rank_{2,2,t2} & \cdots & Rank_{2,10,1} & \cdots & Rank_{2,10,t10} \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots \\
Rank_{u,1,1} & \cdots & Rank_{u,1,t1} & Rank_{u,2,1} & \cdots & Rank_{u,2,t2} & & Rank_{u,10,1} & \cdots & Rank_{u,10,t10}
\end{bmatrix}
\end{array}
$$

FIGURE 10 The array for calculating quadrant values

Various personal value tags could now be regarded as feature tags and used as training samples, and the combined calculations of the classification algorithm and feature selection algorithm were implemented. The training samples were used to calculate the similarity with the data to be classified, and the accuracy of the method was obtained by using tenfold cross-validation. This study used the classification algorithm of supervised learning, which is a learning model that is learned or built from the training data (Cuong, Dinh, & Ho, 2012Meng & Kwok, 2013). This study used the Naïve Bayes, Bayes Net, KNN, J48, and LibSVM classification algorithms in the experimental stage.

In order to increase the accuracy of the classification algorithm, the feature selection algorithm can be used together to delete nonrepresentative features. This study adopted the classification algorithm and feature selection algorithm provided by Waikato Environment for Knowledge Analysis (Weka; Baidu, 2016; Hall et al., 2009) and machine learning software, which is suitable for academic purposes, as the optimum choice of feasibility verification tools.

4 | IMPLEMENTATION AND EXPERIMENT

In a Microsoft Windows 10 operating environment, MySQL, Visual Studio 2015, Xampp, CKIP, and Weka were integrated and combined with the Facebook API to implement a user data collection website to test and verify the proposed method. In the experimental process (Figure 11), the subject clicked on the "Authorization and questionnaire" function to authorize this mechanism to obtain the experimenter's data and test the PVQ scale. Another important data source for this study was hashtag posts derived from Facebook (Figure 12) and Instagram (Figure 13). However, Facebook does not provide an API for extracting hashtag posts, and the API provided by Instagram was inapplicable to this study. Therefore, this study used the webpage of Facebook mobile and Instagram's web viewer website pikore for the tag search to extract the required data.

The improved association rule proposed in this study reduced the number of unwanted itemsets and could therefore shorten the computing time greatly. Figure 14 shows the execution of data on 1,053 items. The *min_Support* was set as 1% of the items; the horizontal axis was the

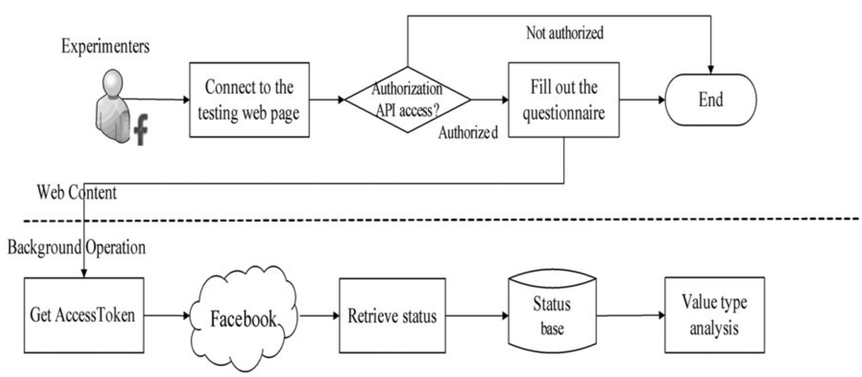

FIGURE 11 The data collection process

FIGURE 12 Hashtag post example from Facebook

FIGURE 13 Hashtag post example from Instagram

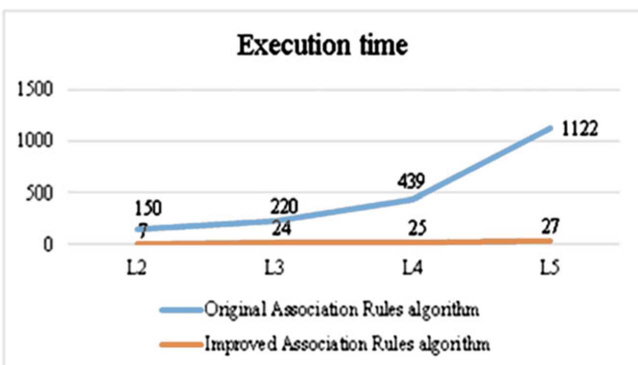

FIGURE 14 Comparison of the improved and original association rule algorithms

frequent-itemsets in lengths of 2, 3, 4, and 5; and the vertical axis was operation time (seconds). It was observed that the computing time was shortened significantly when frequent-itemsets with a length of 2 were generated.

The *min_Support* and *min_Confidence* were set at 4% in this study, so as to reserve the reliable personal value tag. The experiment showed the results deduced from using the tags “commonwealth” and “equality”. For example (Table 1), the “commonwealth” tag contained 1,053 data items, and the “equality” tag contained 916 items. The support of “compassion” deduced from “commonwealth” was 55. The personal value of (support/total number of items) was 5.2%, which was greater than the preset standard. Therefore, “compassion” was classified as the personal value tag of “commonwealth.” Table 1 shows the results of the correlation calculation using SVIs in this study.

There were 61 experiment participants in this study. Their statuses were obtained through the API of Facebook and 20,868 statuses. The TF value of each user's status, term weight table, and Simscore were then calculated. Each participant could obtain the corresponding weights (Figure 15) that were extracted from the partial result of the Simscore computation of the experimenters.

TABLE 1 Result of association rule computing of all SVIs (#Chinese [English])

Specific personal value item (SVI)	Result	Total number of item (support)
#公益 (Commonweal)	#愛心 (Compassion)	55 (5.2%)
#世界和平(A world at peace)	#身體健康 (Healthy body)	42 (4.2%)
#平等 (Equality)	#自由 (Freedom)	56 (6.1%)
#安全 (Family security)	#安心 (Relieved)	466 (16%)
#成功 (Successful)	#人生 (Life)	121 (4%)
#自由 (Freedom)	#旅行 (Travel)	46 (4%)
#自律 (Self-discipline)	—	—
#孝順 (Honouring of parents and elders)	#父母 (Parents)	39 (4%)
#享受生活 (Enjoyed life)	#享受 (Enjoy)	60 (4.2%)
#忠誠 (Obedient)	#團結 (Unity)	29 (4.1%)
#放縱 (Self-indulgent)	#放鬆 (Relax)	20 (5.2%)
#勇氣 (Daring)	#夢想 (Dream) #自信 (Confidence) #希望 (Hope) #愛 (Affection)	372 (9%) 399 (10%) 237 (6%) 331 (8.6%)
#負責 (Responsible)	#責任 (Responsibility)	24 (4.1%)
#面子 (Preserving my public image)	—	—
#能力 (Capable)	#努力 (Strive)	36 (4.2%)
#虔誠 (Devout)	#信仰 (Faith)	113 (12.8%)
#財富 (Wealth)	#創業 (Venture) #健康 (Health) #愛情 (Love)	164 (14%) 284 (25%) 144 (12.8%)
#乾淨 (Clean)	—	—
#創造力 (Creativity)	#想像力 (Imagination)	126 (6.25%)
#愉悅 (Pleasure)	#開心 (To be joyful) #放鬆 (Relax)	140 (10.1%) 71 (5%)
#智慧 (Wisdom)	#人生 (Life)	162 (14%)
#節制 (Moderate)	#運動 (Sport)	22 (4.4%)
#誠實 (Honest)	#面對 (Face) #信任 (Trust)	37 (4%) 37 (4%)
#認命 (Accept my portion in life)	#無奈 (Helpless)	24 (4.8%)
#寬容 (Forgiving)	#快樂 (Happy) #包容 (Inclusive)	25 (4.3%) 25 (4.3%)
#影響力 (Influential)	—	—
#獨立 (Independent)	#民主 (Democracy)	56 (4.1%)
#環境保護 (Protecting the environment)	#大自然 (Nature) #生態 (Ecology) #環境 (Environment)	40 (4.1%) 62 (6.4%) 40 (4.1%)
#謙卑 (Humble)	#感恩 (Thanksgiving)	29 (4%)
#禮貌 (Politeness)	#尊重 (Respect) #態度 (Attitude)	65 (7.8%) 47 (5.6%)
#權力 (Authority)	#人緣 (Popularity)	34 (4.9%)
#傳統 (Respect for tradition)	—	—

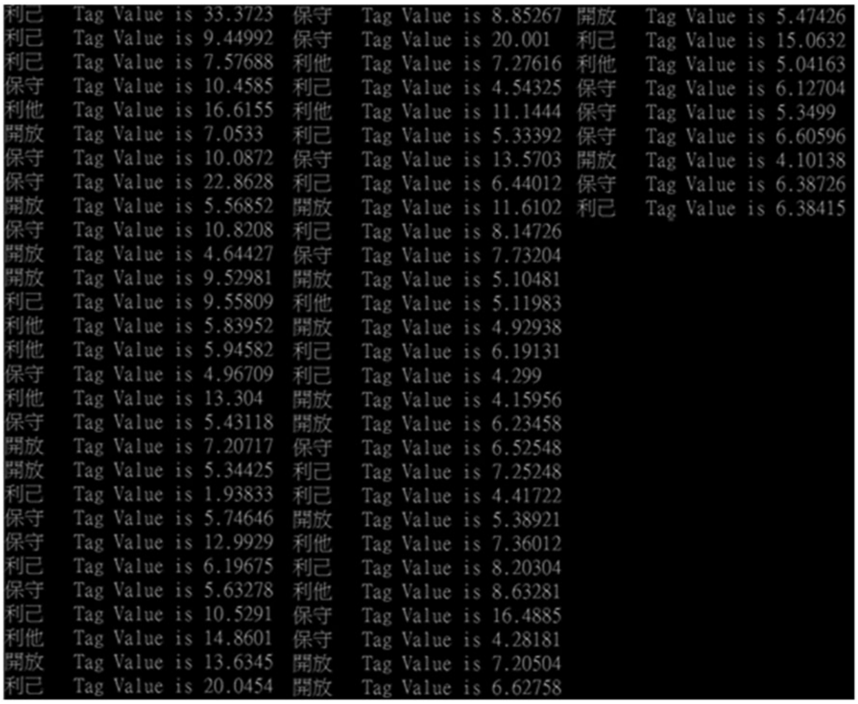

FIGURE 15 Partial result of Simscore computation

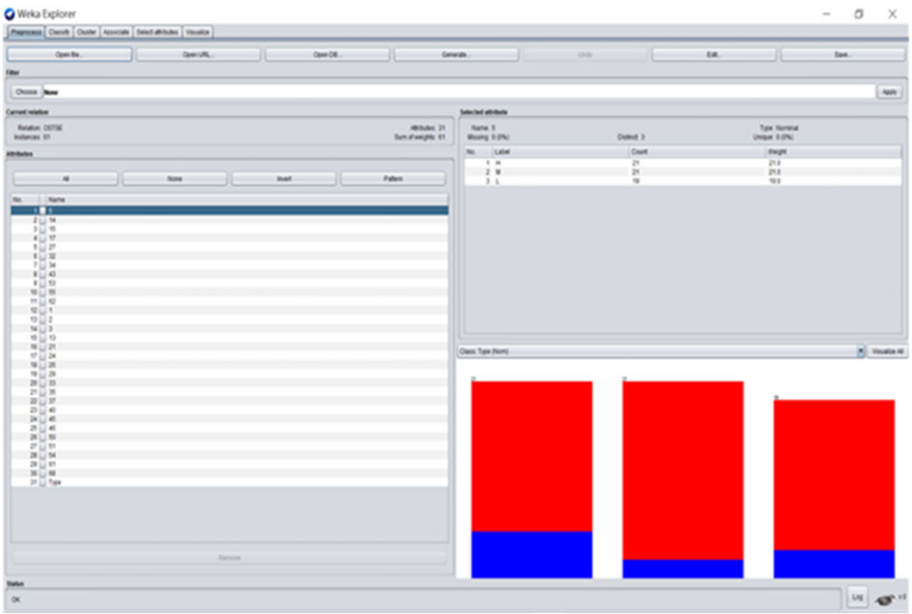

FIGURE 16 Result of classification analysis using Weka

Based on the tags, the scores were next divided into high, medium, and low groups. The personal value tags of various types of values were used to analyse the 61 participants' personal value orientation. The results of this analysis are only illustrated with short tables in this paper because the completed results were long and would be difficult to reproduce. The simplified results are shown in Appendix A.

5 | CONCLUSION AND DISCUSSION

5.1 | Discussion of the key findings

The above data were analysed using Weka data analysis software (Figure 16). Tenfold cross-validation was combined with the classification algorithm and feature selection algorithm for calculation. The results are shown in Appendix B.

The analysis results showed precision without the feature selection algorithm. LibSVM performed better universally. When the feature selection method CfsSubset Eval was used, the optimum accuracy in self-transcendence and self-enhancement was 83.6%, the optimum precision in openness to change and conservation was 73.7%, and the optimum precision in analysing quadrant values was 70.4%. The experimental results proved that the precision of this method in analysing the Facebook value type was much better than the randomly analysed value. This analysis method was therefore feasible.

5.2 | Summary and conclusion

This study developed an automatic personal value analysis method that could predict the Schwartz value type by analysing Facebook user data. This study used hashtag posts from Facebook and Instagram as the training data, and the value type was estimated by analysing the status of Facebook users.

In order to determine the features of each value, this study used Schwartz SVIs (Sagiv & Schwartz, 2000) as seeds. Many tag sets were extracted from the hashtag posts disclosed in Facebook and Instagram and then analysed using the improved association rule algorithm to identify the personal value tags, which were used as the features. It was found after implementation that the improved association rule algorithm shortened the computing time and preserved the required data. The *min_Support* and *min_Confidence* were set at 4% in this study. Some effective tags were filtered out, but most of the insignificant tags were also filtered out. The tags deduced by the algorithm were correlated with SVI to some extent.

This study designed a consistency model and relativity model that were combined with *aTFIDF* to calculate the weights for hashtag posts. It was found after implementation that the weights of the terms correlated with the personal value tag were enhanced. Finally, for analysing the personal value type, this study employed the Simscore of the personal value tag for calculation; however, the personal value of some tags could not be distinguished clearly, leading to poor accuracy. Therefore, this study divided the Simscore values into high, medium, and low levels based on the tag and used the feature selection algorithm and classification algorithm for analysis.

According to the analysis results, in cases without the feature selection algorithm, the classification algorithm LibSVM had the optimum accuracy universally. In cases using the feature selection algorithm CfsSubset Eval, LibSVM still maintained the optimum accuracy in self-transcendence and self-enhancement, and the precision of Naïve Bayes, Bayes Net, and 5-IBk (knn) in openness to change and conservation increased to 73.7%. The accuracy of Naïve Bayes and Bayes Net in analysing quadrant values increased to 70.4%. According to the experiment, the accuracy significantly rose after feature screening, and the precision was higher than 70%, proving that the precision was reliable.

Previous studies on personal values were only aimed at discussions and applications on questionnaires, and they seldom implemented automated analysis of Facebook users' information. The experimental results showed that the proposed method for personal value type analysis was feasible for judging Facebook users' personal values.

During the extraction of hashtag posts in the experimental process, although there were numerous hashtag posts in social media, when the SVI of this study was used to extract hashtag posts, the number either varied or was too small. Therefore, the accuracy could be influenced. This study used a user's status as the basis of judgement. When the status information published by the user was not complete, there might be poor results.

Most previous research was based on the data of social media users, which may result in data limitations. The training data used by this study was based on the topic tags of public community media. The dynamic articles used by social media were labelled; therefore, the advantages of this study included the following:

1. This study used public community media topic tags as articles. The data source was broad and did not only consider the data of social media users. Therefore, the source of the training data was an advantage.
2. Because the training data used information from open social media, which varies with time, the training data will not be representative because time is too long.
3. This study used association rules to expand the value tags. In practice, it could be used to find a value tag that is more in line with the time.

This study had a number of limitations. The suggestions for follow-up research were as follows:

1. This study only discussed the statuses and tags of hashtag posts in social media, whereas hashtag posts often involve picture and video information. In order to probe into these contents, further techniques are required, such as image analysis techniques.
2. When SVIs are used to extract hashtag posts, the obtained tags may be a compound of SVI. For example, when the SVI of "pleasure" is used to collect hashtag posts, hashtag posts for "a pleasant day" may be obtained. This type of problem was screened manually in this study. In order to increase efficiency, automated screening methods should be developed in the future.

ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Science and Technology of the Republic of China, Taiwan, for financially supporting this research under Contracts MOST 103-2221-E-343-003-MY3 and MOST 107-2221-E-343-003-MY3.

ORCID

Tsung-Yi Chen https://orcid.org/0000-0002-6850-0680

REFERENCES

Adedoyin-Olowe, M., Gaber, M. M., Dancausa, C. M., Stahl, F., & Gomes, J. B. (2016). A rule dynamics approach to event detection in Twitter with its application to sports and politics. *Expert Systems with Applications*, *55*, 351–360. https://doi.org/10.1016/j.eswa.2016.02.028

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, VLDB (Vol. 1215) (pp. 487-499).

Baidu. (2016). Weka. http://baike.baidu.com/item/weka.

Bhargava, P., Brdiczka, O., & Roberts, M. (2015). Unsupervised modeling of users' interests from their Facebook profiles and activities. In proceedings of the 20th international conference on intelligent user Interface, 191–201.

Chen, J., Hsieh, G., Mahmud, J. U., & Nichols, J. (2014). Understanding individuals' personal values from social media word use. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, 405-414.

Chen, K. J., & Ma, W. Y. (2005). Design of CKIP Chinese word segmentation system. *Chinese and Oriental Languages Information Processing Society*, *14*(3), 235–249.

Cuong, N. V., Dinh, V., & Ho, L. S. T. (2012). Mel-frequency cepstral coefficients for eye movement identification. In 2012 IEEE 24th international conference on tools with artificial intelligence, pp. 253–260.

ElTayeby, O., Molnar, P., & George, R. (2014). Measuring the influence of mass media on opinion segregation through Twitter. *Procedia Computer Science*, *36*, 152–159. https://doi.org/10.1016/j.procs.2014.09.062

Feng, F., Cho, J., Pedrycz, W., Fujita, H., & Herawan, T. (2016). Soft set based association rule mining. *Knowledge-Based Systems*, *111*, 268–282. https://doi.org/10.1016/j.knosys.2016.08.020

Flodén, J., & Williamsson, J. (2016). Business models for sustainable biofuel transport: The potential for intermodal transport. *Journal of Cleaner Production*, *113*, 426–437. https://doi.org/10.1016/j.jclepro.2015.11.076

Fritscher, B., & Pigneur, Y. (2009, September). *Supporting business model modelling: A compromise between creativity and constraints. In International Workshop on Task Models and Diagrams for User Interface Design* (pp. 28–43)Springer Berlin Heidelberg.

Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. In CHI'11 extended abstracts on human factors in computing systems. pp. 253-262.

Gou, L., Zhou, M. X., & Yang, H. (2014). KnowMe and ShareMe: Understanding automatically discovered personality traits from social media and user sharing preferences. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 955–964.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*(Mar), 1157–1182.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18. https://doi.org/10.1145/1656274.1656278

Hamed, A. A., Wu, X., Erickson, R., & Fandy, T. (2015). Twitter KH networks in action: Advancing biomedical literature for drug search. *Journal of Biomedical Informatics*, *56*, 157–168. https://doi.org/10.1016/j.jbi.2015.05.015

Hofstede, G. (1980). *Cultures consequences: International differences in work-related values*. Beverly Hills, CA: Sage.

Jabeur, N., Nait-Sidi-Moh, A., & Zeadally, S. (2017). Crowd social media computing: Applying crowd computing techniques to social media. *Applied Soft Computing*, *66*, 495–505.

Kaplan, S. (2012). *The business model innovation factory: How to stay relevant when the world is changing*John Wiley & Sons. https://doi.org/10.1002/9781119205234

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Krystallis, A., Vassallo, M., & Chryssohoidis, G. (2012). The usefulness of Schwartz's 'Values Theory' in understanding consumer behaviour towards differentiated products. *Journal of Marketing Management*, *28*(11–12), 1438–1463. https://doi.org/10.1080/0267257X.2012.715091

Linli, (2016). Facebook user number Asia increase, 90 million more than a year. TechNews. Retrieved from http://technews.tw/2016/03/16/facebook-sees-big-growth-in-asia-despite-free-basics-controversy/.

Meng, Y., & Kwok, L. F. (2013). Enhancing false alarm reduction using voted ensemble selection in intrusion detection. *International Journal of Computational Intelligence Systems*, *6*(4), 626–638. https://doi.org/10.1080/18756891.2013.802114

Ojasalo, J., & Ojasalo, K. (2015). Service logic business model canvas for lean development of SMEs and start-ups. Handbook of Research on Entrepreneurship in the Contemporary Knowledge-Based Global Economy, 217.

Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, *31*, 527–541. https://doi.org/10.1016/j.chb.2013.05.024

Rokeach, M. (1973). *The nature of human values* (Vol. 438). New York: Free press.

Sagiv, L., & Schwartz, S. H. (2000). Value priorities and subjective well-being: Direct relations and congruity effects. *European Journal of Social Psychology*, *30*(2), 177–198. https://doi.org/10.1002/(SICI)1099-0992(200003/04)30:2<177::AID-EJSP982>3.0.CO;2-Z

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, *24*(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

Sawant, V., & Shah, K. (2016). Performance evaluation of distributed association rule mining algorithms. *Procedia Computer Science*, *79*, 127–134. https://doi.org/10.1016/j.procs.2016.03.017

Schein, E. H. (1985). *Organizational culture and leadership Jossey-Bass*. San Francisco.

Schwandt, T. A. (1994). Are there universal aspects in the content and structure of values. *Journal of Social Sciences*, *50*(1), 19–45.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, *25*, 1–65.

Schwartz, S. H., & Sagiv, L. (1995). Identifying culture-specifics in the content and structure of values. *Journal of Cross-Cultural Psychology*, *26*(1), 92–116. https://doi.org/10.1177/0022022195261007

Skowron, M., Tkalčič, M., Ferwerda, B., & Schedl, M. (2016). Fusing social media cues: Personality prediction from Twitter and Instagram. In proceedings of the 25th international conference companion on world wide web, 107–108.

Small, T. A. (2011). What the hashtag? A content analysis of Canadian politics on Twitter. *Information, Communication & Society*, *14*(6), 872–895. https://doi.org/10.1080/1369118X.2011.554572

Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics—Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, *39*, 156–168. https://doi.org/10.1016/j.ijinfomgt.2017.12.002

Swartz, S. (1996). *Value priorities and behavior: Applying a theory of integrated value systems in the psychology of values*. In C. Seligman, J. M. Olson, & M. P. Zanna (Eds.). Mahwah, NJ: Lawrence Erlbaum. Retrieved from http://www.questia.com/PM.qst

Tajudeen, F. P., Jaafar, N. I., & Ainin, S. (2018). Understanding the impact of social media usage among organizations. *Information and Management*, *55*(3), 308–321.

Tian, J. M. (2008). *The impacts of value and environmental cognition on pro-environmental behaviors—Taipei metropolitan area case study*. Taipei: National Taipei University.

Toro-Jarrín, M. A., Ponce-Jaramillo, I. E., & Güemes-Castorena, D. (2016). Methodology for the of building process integration of Business Model Canvas and Technological Roadmap. *Technological Forecasting and Social Change*, *110*, 213–225. https://doi.org/10.1016/j.techfore.2016.01.009

Wang, C. (2012). *Validation of the structure of Schwartz human value theory: An empirical verification in Taiwan*. Kaohsiung: National Kaohsiung Normal University.

Wei, H., Zhang, F., Yuan, N. J., Cao, C., Fu, H., Xie, X., & Ma, W. Y. (2017). Beyond the words: Predicting user personality from heterogeneous information. In proceedings of the tenth ACM international conference on web search and data mining, 305–314.

Wikipedia. (2016a). Social Media. https://en.wikipedia.org/wiki/Social_media.

Wikipedia, (2016b), Value. http://en.wikipedia.org/wiki/Social_media#Distinction_from_other_media.

Yang, C., Pan, S., Mahmud, J., Yang, H., & Srinivasan, P. (2015). Using personal traits for brand preference prediction. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, 86-96.

Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, *112*(4), 1036–1040. https://doi.org/10.1073/pnas.1418680112

Zhang, Y., & Pennacchiotti, M. (2013). Predicting purchase behaviors from social media. In Proceedings of the 22nd international conference on World Wide Web, 1521-1532.

Zhao, F., Zhu, Y., Jin, H., & Yang, L. T. (2015). A personalized hashtag recommendation approach using LDA-based topic model in microblog environment. *Future Generation Computer Systems*, *65*, 196–206.

AUTHOR BIOGRAPHIES

Dr. Tsung-Yi Chen is currently a Professor of Department of Information Management, and Institute of Information Management, College of Science and Technology, Nanhua University, Taiwan, ROC. He gained his PhD degrees from Institute of Manufacturing Engineering, National Cheng Kung University, Taiwan, ROC, in 2006. His research interests include Enterprise System and Information Integration, Knowledge Access Control and Sharing for Virtual Enterprises, Security in Social Network and Virtual Community, Electronic Commerce Business Model Innovation, Advanced Network Marketing Technology, and Cloud-based Enterprise Information System. He has published his research results in various journals.

Mr. Hsiang-An Cheng is currently a data scientist at INNOLUX Corporation. He received his master degree from Institute of Manufacturing Information and System, National Cheng Kung University in 2016. His research specialties are data analysis and text mining.

Dr. Yuh-Min Chen is currently the Distinguished Professor of Department of Computer Science and Information Engineering, and Institute of Manufacturing Information and Systems, College of Electric Engineering and Computer Science, National Cheng Kung University, Taiwan, ROC. His research has been focused on the applications of knowledge engineering and management in virtual enterprising, including areas of Collaborative Engineering Knowledge Management, Semantic Content Management, Semantic-based Image Management, Web-based Business Intelligence, and KM (Knowledge Management)-based e-Learning.

How to cite this article: Chen T-Y, Cheng H-A, Chen Y-M. Developing a personal value analysis method of social media to support customer segmentation and business model innovation. *Expert Systems*. 2019;36:e12374. https://doi.org/10.1111/exsy.12374

APPENDIX A

TABLE A1 Training result of self-transcendence and self-enhancement

Experimenter Personal value tag	1	2	3	4	5	...	61
$Tag_{1,\ 1}$	H	M	L	M	H	...	L
$Tag_{1,\ 2}$	H	M	M	H	M	...	H
$Tag_{1,\ 3}$	M	L	H	L	M	...	M
$Tag_{1,\ 4}$	H	H	L	H	H	...	M
$Tag_{1,\ 5}$	H	M	H	H	H	...	H
$Tag_{1,\ 6}$	H	H	H	M	H	...	L
⋮	⋮	⋮	⋮	⋮	⋮	⋱	
$Tag_{4,\ t4}$	L	M	H	M	M	...	M
Type	SE	ST	SE	ST	ST	...	ST

TABLE A2 Training result of openness to change and conservation

Experimenter Personal value tag	1	2	3	4	5	...	61
$Tag_{5,\ 1}$	H	L	M	H	H	...	M
$Tag_{5,\ 2}$	H	M	H	M	H	...	L
$Tag_{5,\ 3}$	H	M	L	H	H	...	H
$Tag_{5,\ 4}$	H	M	L	M	H	...	L
$Tag_{5,\ 5}$	M	H	L	M	L	...	M
$Tag_{5,\ 6}$	H	L	H	M	H	...	H
⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮
$Tag_{10,\ t10}$	L	M	H	M	H	...	M
Type	CO	OP	OP	CO	CO	...	OP

TABLE A3 Training result of quadrant values

Experimenter Personal value tag	1	2	3	4	5	...	61
$Tag_{1,\ 1}$	H	M	L	M	H	...	L
$Tag_{1,\ 2}$	H	M	M	H	M	...	H
$Tag_{1,\ 3}$	M	L	H	L	M	...	M
$Tag_{1,\ 4}$	H	H	L	H	H	...	M
$Tag_{1,\ 5}$	H	M	H	H	H	...	H
$Tag_{1,\ 6}$	H	H	H	M	H	...	L
⋮	⋮	⋮	⋮	⋮	⋮	⋱	⋮
$Tag_{10,\ t10}$	L	M	H	M	H	...	M
Type	STOP	STOP	SEOP	STCO	STCO	...	STOP

APPENDIX B

TABLE B1 Accuracy of self-transcendence and self-enhancement

		Feature selection algorithm	
		No using (%)	CfsSubset eval (%)
Classification algorithms	Naïve Bayes	67.2	78.6
	Bayes Net	68.8	78.6
	1-IBk (knn)	72.1	78.6
	3-IBk (knn)	73.7	77.0
	5-IBk (knn)	78.6	77.0
	J48 (DT)	83.6	83.6
	LibSVM	83.6	83.6

TABLE B2 Accuracy of openness to change and conservation

		Feature selection algorithm	
		No using (%)	CfsSubset eval (%)
Classification algorithms	Naïve Bayes	50.8	73.7
	Bayes Net	54.0	73.7
	1-IBk (knn)	70.4	70.4
	3-IBk (knn)	60.6	70.4
	5-IBk (knn)	70.4	73.7
	J48 (DT)	49.1	63.9
	LibSVM	72.1	72.1

TABLE B3 Accuracy of quadrant values

		Feature selection algorithm	
		No using (%)	CfsSubset eval (%)
Classification algorithms	Naïve Bayes	40.9	70.4
	Bayes Net	44.2	70.4
	1-IBk (knn)	44.2	55.7
	3-IBk (knn)	55.7	62.2
	5-IBk (knn)	59.0	60.6
	J48 (DT)	45.9	47.5
	LibSVM	62.2	62.2