
Shen-Tsu Wang¹

¹Department of Commerce Automation and Management, National Pingtung University Taiwan, Province of China

Abstract

While it is difficult to avoid uncertainties when shopping on the Internet, trust can reduce customers’ perceived uncertainties, and enhance their willingness and frequency to buy products and services. The difference in time and space information transparency between customers and on-line sellers, as well as the complex unpredictability of network structure, result in frequent uncertainty for on-line transactions. Therefore, through text mining and integrating the Genetic Algorithm (GA) with the Support Vector Machine (SVM), this project classifies the data of on-line group buying community complaints according to the posts left on Facebook and the three major group-buying websites of Taiwan. The terms are selected based on term frequency, document frequency, uniformity, and conformity, while document classification effectiveness is calculated using precision, recall rate, and F-measure. Community complaints are classified into the uncertain performance indicators that influence on-line group buying for integrated statistics, in order that specific performance indicators of community group-buying websites can be generated. Afterwards, based on the on-line group buying community performance indicator sequence, as integrated according to the dynamic Multicriteria Optimization and Compromise Solution (VIKOR) method and prosperity countermeasure signals, grey correlation sorting is applied to analyze the dynamic performance indicator sequence of different communities, in order to determine the life context of different populations for the reference of on-line group buying providers.

Keywords: Genetic algorithm, Support vector machine, On-line group buying, Life context, Grey correlation sorting, Dynamic performance indicator.

Contents

1. Introduction ........................................................................................................................................... 174
2. Research Method .................................................................................................................................... 174
3. Results Analysis ..................................................................................................................................... 180
4. Conclusion ........................................................................................................................................... 180
References .................................................................................................................................................. 180


DOI: 10.20448/journal.500/16.3.3/500.3.173.181

ISSN(E) : 2313-7401
ISSN(P) : 2518-0096
Licensed: This work is licensed under a Creative Commons Attribution 3.0 License.

Funding: This study received no specific financial support.
Competing Interests: The author declares that there are no conflicts of interests regarding the publication of this paper.
Transparency: The author confirms that the manuscript is an honest, accurate, and transparent account of the study was reported; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained.
Ethical: This study follows all ethical practices during writing.

History: Received: 3 February 2016 / Revised: 25 February 2016 / Accepted: 29 February 2016 / Published: 2 March 2016
Publisher: Asian Online Journal Publishing Group
1. Introduction

The course of the real world and contexts of places of existence are the details of human life. The theory of humanistic geography opposes abstract objective and mechanized intuitionistic humanity, and places stress on the relative meanings, different values, and objectives of persons throughout the world. Since the theory of humanistic geography pays great attention to the context of humanistic life, it is attracted and used by the theoretical concept of phenomenology, and its interesting subjects are examined, such as correlated place, mutual space, and life area (Wen, 2005). Scholars who think based on context not only consider the single development of users and the actual nature of products, they also emphasize interactions, constructed cultural aspects, and process values, and pay attention to the common structural relationship between multilingual sociocultural values and technology (Dulk et al., 2013; Trefalt et al., 2013). The uncertainties of life context include the constraints of innovation sources, as well as the differences between manufacturing and customer sides. The important findings regarding the constraints of innovation sources show that, the more accustomed individuals are to observing or using a tool in a particular way, the harder it is to escape the existing constraint architecture, and the harder it is to have innovative views of substantial content, thus, research processes are unlikely to propose novel views (Hippel, 2011).

While it is difficult to avoid uncertainties when shopping on the Internet, trust can reduce customers’ perceived uncertainties, and enhance their willingness and frequency to buy products and services. The difference in time and space information transparency between customers and on-line sellers, as well as the complex unpredictability of network structure, result in frequent uncertainty for on-line transactions. Therefore, through text mining and integrating the Genetic Algorithm (GA) with the Support Vector Machine (SVM), this project classifies the data of on-line group buying community complaints according to the posts left on Facebook and the three major group-buying websites of Taiwan. The terms are selected based on term frequency, document frequency, uniformity, and conformity, while document classification effectiveness is calculated using precision, recall rate, and F-measure. Community complaints are classified into the uncertain performance indicators that influence on-line group buying for integrated statistics, in order that specific performance indicators of community group-buying websites can be generated. Afterwards, based on the on-line group buying community performance indicator sequence, as integrated according to the dynamic Multicriteria Optimization and Compromise Solution (VIKOR) method and prosperity countermeasure signals, grey correlation sorting is applied to analyze the dynamic performance indicator sequence of different communities, in order to determine the life context of different populations for the reference of on-line group buying providers (Chen et al., 2010; Cheng and Huang, 2013; Wang et al., 2013; Liang et al., 2014; Zhou and Xie, 2014; Hsu et al., 2015).

2. Research Method

This project uses the Chinese Knowledge Information Processing (CKIP) of Academia Sinica to analyze the word segmentation of Chinese documents, and screens out the terms of on-line group buying community complaints. Afterwards, the optimal term combination is selected by GA to train the SVM of the existing types of documents, and the classificatory documents are classified by test file; then all other documents not classified into the existing types are clustered, and document clustering is optimized using GA (Lo et al., 2008; Trstenjak et al., 2014; Erra et al., 2015). There are two major steps, as follows.

2.1. Document Pre-Processing

The document classification of this project can be divided into two parts, the original type of document cluster and new document cluster. The existing type of document cluster contains the existing architecture type of original documents, while the new document cluster contains partial existing type of documents, as well as any documents not belonging to the original type. As there is no appropriate decision-making approach for terms in Chinese documents, word segmentation is required in order to recognize all terms in the documents. Among the plans for Chinese word segmentation, as constructed by the Chinese Knowledge Information Processing Group (CKIPG) of Academia Sinica, CKIP is representative and practical. Therefore, this project uses CKIP, as provided by CKIPG, to process the training documents and test documents of this project.

2.2. GA-SVM Model

This project combines GA with SVM to construct the GA-SVM computational model, in order to determine better term combinations to train the existing type of classification model, and whether or not new documents can be classified into the original type is judged according to this architecture. In order to effectively determine better term combinations, important terms are selected using GA, where the Term Frequency-Inverse Document Frequency (TF-IDF) values are calculated according to the selected terms (Lo et al., 2008; Trstenjak et al., 2014; Erra et al., 2015) as the basis of training documents and create the vector values of the documents, and the obtained document vectors are used as the input vector values to train the GA-SVM model base. The term combinations, as selected by this model, correspond to the contents of new documents in order to calculate the TF-IDF-based document vector matrices of the new documents, and the document vector matrices of the new documents are converted into the SVM format and imported into the SVM training model for classification. When new documents fail to be classified as the original type, they are extracted for second stage processing. However, when GA is selected, as it can implement efficient global search for a better solution to the problem, which can reduce the probability of obtaining a local optimal. Where TF-IDF is expressed as Eq. (1).

\[
TF - IDF(W_i,M) = TF(W_i,M) \times IDF(W_i) = TF(W_i,M) \times \log \left( \frac{|P|}{DF(W_i)} \right)
\]
Where,

\( TF - IDF (W_i, M) \): Weight of term \( W_i \) in \( M \)

\( |D| \): Total number of messages in document

\( DF (W_i) \): Occurrence frequency of term \( W_i \) in document set

The message can represent the matrix \( M \) formed by the vector of a term, expressed as Eq. (2).

\[
M = \begin{bmatrix}
  w_{11} & w_{12} & \cdots & w_{1p} \\
  w_{21} & w_{22} & \cdots & w_{2p} \\
  \vdots & \vdots & \ddots & \vdots \\
  w_{m1} & w_{m2} & \cdots & w_{mp}
\end{bmatrix}
\]  

(2)

where,

\( W_a \): Weight of term \( k \) in document \( i \) \((i = 1...m)\)

\( m = |D| \)

\( k = 1,...,p \)

\( p \): Number of terms extracted from all messages

Therefore, this project intends to use the characteristics of GA to determine the most representative terms to optimize the SVM model, in order to classify the complaints of on-line group-buying communities. The execution flow of the GA-SVM classifier is introduced below. This project uses the four functions of Chou et al. (2007) as the basis for selecting terms, the term frequency, document frequency, uniformity, and conformity, described as follows:

1) Select term
(1) Term frequency (TF)

\( TF \) represents the term occurrence probability of term \( i \) in \( K_j \) type, expressed as Eqs. (3)-(4):

\[
TF_{ij} = \frac{T F_{ij}}{\sum_{j=1}^{J} T F_{ij}}
\]  

(3)

\[
T F_{ij} = \frac{a_{ij}}{\sum_{j=1}^{J} a_{ij}}
\]  

(4)

where,

\( i \): Term 

\( j \): Type \( K_j \)

\( a_{ij} \): Occurrence number of term \( i \) in \( K_j \) type

\( T F_{ij} \): Occurrence probability of term \( i \) in \( K_j \) type

If the occurrence number of a term \( i \) in \( K_j \) type is greater than that in other types, and the calculated TF is large; this term \( i \) can represent type \( K_j \).

(2) Document frequency (DF)

\( DF \) represents the occurrence probability in documents of term \( i \) in \( K_j \) type, expressed as Eqs. (5)-(6):

\[
DF_{ij} = \frac{D F_{ij}}{\sum_{j=1}^{J} D F_{ij}}
\]  

(5)

\[
D F_{ij} = \frac{S_i}{S_j}
\]  

(6)

where,

\( S_i \): Occurrence number of documents with term \( i \) in \( K_j \) type

\( D F_{ij} \): Occurrence probability of documents with term \( i \) in \( K_j \) type

If the number of documents with term \( i \) in \( K_j \) type is greater than the number of documents in other types, then term \( i \) better represents \( K_j \) type.

(3) Uniformity

Uniformity represents the occurrence probability of term \( i \) in all documents of \( K_j \) type, expressed as Eqs. (7)-(8):

\[
U_{ij} = -\sum_{x=1}^{S_i} h_{ix} \log_h x
\]  

(7)
\[ b_i = \frac{f_{i,c}}{\sum_{j=1}^{S_j} f_{i,j}} \]  

where,

\( c \): Document

\( S_j \): Total number of documents within \( K_j \) type

\( f_{i,c} \): Occurrence number of term \( i \) in \( c \) documents

\( b_i \): Average occurrence probability of term \( i \) in document \( c \)

If the occurrence number of term \( i \) in \( K_j \) type documents is greater than the occurrence number of other terms in \( K_j \) type documents, the larger the calculated uniformity, the better this term \( i \) represents this type than other terms.

(4) Conformity

The conformity represents the probability of the occurrence of term \( i \) in all types, expressed as Eqs. (9)–(10):

\[ CF_i = -\sum_{j=1}^{J} P_{ij} \log P_{ij} \]  

\[ P_{ij} = \frac{S_{ij}}{\sum_{j=1}^{S_j} S_{ij}} \]  

where,

\( P_{ij} \): Probability of occurrence of document \( S \) with term \( i \) in \( j \) type

When term \( i \) occurs in all types, the \( CF_i \) value is large. Contrarily, when \( i \) only occurs in one single type, the \( CF_i \) value is 0. Therefore, the smaller the \( CF_i \), the better the term \( i \) represents the type.

2) Classifier training by SVM

Each term will obtain four threshold values. This project selects the terms meeting the four thresholds as the basis of training documents, then the TF-IDF value of the obtained term is calculated, and the document vector value is created for each document as the input vector of SVM for the training classifier.

(1) Model parameter selection

This project uses a library for support vector machines (LIBSVM) (Chang and Lin, 2001) for document classification, as LIBSVM provides multiple parameter settings, four core functions, four classifiers, and an important parameter search tool, thus, enabling the user to adjust parameters according to different problems in order to increase the efficiency of classification. However, as document classification is mostly a polytype classification problem, this project uses the C-Support Vector Classification (C-SVC) form, which supports the multi-classification model to train SVM. The terms meeting the threshold set value are extracted, the document vector is created, and the vector value is imported into LIBSVM to train the SVM classifier.

(2) Predict training data by SVM

When the SVM model of some chromosomes is generated, the SVM classification model can be used to predict the classification of training documents. As this project uses GA for continuous calculation to determine the representative term combination, the condition of calculation is to calculate the fitness value according to the mean F-measure of SVM. Afterwards, if the calculation reaches the stop condition, and the SVM has good classification effectiveness, we can use this classification model to classify new community complaint documents.

3) GA-SVM fitness function design

In the course of document classification, document classification effectiveness is calculated based on Precision, Recall, and F-measure. The fitness function of this project uses the F-measure mean as the basis of evaluating classification. Therefore, the larger the fitness function value, the better the extracted term classifies the document into the original type, where the fitness function is expressed as Eqs. (11)–(14):

\[ fitness = \frac{\sum_{i=1}^{J} F_{s,p,K_i}}{I} \]  

F-measure: 

\[ F_{s,p,K_i} = \frac{2P_{s,p,K_i} \times R_{s,p,K_i}}{P_{s,p,K_i} + R_{s,p,K_i}} \]  

Precision: 

\[ P_{s,p,K_i} = \frac{N_{p,K_i} \cap N_{s,K_i}}{N_{s,K_i}} \]  

Recall: 

\[ R_{s,p,K_i} = \frac{N_{p,K_i} \cap N_{s,K_i}}{N_{p,K_i}} \]  

where,

\( i \): type \( K_i \)
\( N_{s,K} \): Number of documents classified as \( K \) by classifier
\( N_{p,K} \): Number of documents defined as \( K \) by training document
\( N_{p,K} \cap N_{s,K} \): Number of documents classified as \( K \) by training document and classifier

The aforesaid analyzed on-line group buying complaints are put in the classification influencing on-line group buying community uncertainty in Table 1 and calculated.

<table>
<thead>
<tr>
<th>Group-buying website performance aspect</th>
<th>Group-buying website performance indicator</th>
<th>Operational definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Community word-of-mouth acceptance</td>
<td>A1. Consumption experience</td>
<td>Community experience in buying or using the goods.</td>
</tr>
<tr>
<td></td>
<td>A2. Specialty</td>
<td>Community word-of-mouth acceptor's awareness of the merchandise information.</td>
</tr>
<tr>
<td></td>
<td>B1. Positive/negative word-of-mouth</td>
<td>Positive/negative word-of-mouth type of commodity to be bought by community, including fully positive word-of-mouth, fully negative word-of-mouth, and half and half positive/negative word-of-mouth.</td>
</tr>
<tr>
<td></td>
<td>B2. Information quantity</td>
<td>Number of word-of-mouth comments on the commodity to be bought by community.</td>
</tr>
<tr>
<td>B. Community word-of-mouth content</td>
<td>C1. Reputation and evaluation</td>
<td>Whether the past evaluation of the master buyer is good, and whether the master buyer is reliable.</td>
</tr>
<tr>
<td></td>
<td>C2. Friendly interaction</td>
<td>Master buyer responds to members' questions, and informs members of the matters concerned with group buying.</td>
</tr>
<tr>
<td>C. Master buyer reliability</td>
<td>D1. Financial risk</td>
<td>Probable monetary loss of community from on-line shopping.</td>
</tr>
<tr>
<td></td>
<td>D2. Social Risk</td>
<td>Whether the community's shopping on the website is objected by peers.</td>
</tr>
<tr>
<td></td>
<td>D3. Psychological risk</td>
<td>Community has negative opinion after shopping or repents buying.</td>
</tr>
<tr>
<td></td>
<td>D5. Performance risk</td>
<td>Whether the function of a commodity bought from the internet meets the original expectations of the community.</td>
</tr>
<tr>
<td></td>
<td>D6. Time risk</td>
<td>Time spent searching for the desired commodity online and obtaining the commodity.</td>
</tr>
<tr>
<td>D. Community perceived risk</td>
<td>D7. Privacy risk</td>
<td>Identity data registered for on-line transactions are illegally used by others.</td>
</tr>
<tr>
<td></td>
<td>E1. Disturbance-free shopping environment</td>
<td>Take part in on-line group buying for disturbance-free shopping environment.</td>
</tr>
<tr>
<td></td>
<td>E2. Trust group-buying website</td>
<td>Take part in on-line group buying for trusting master buyer or group-buying website.</td>
</tr>
<tr>
<td></td>
<td>E3. Group-buying products are relatively quality</td>
<td>Take part in on-line group buying, as group-buying commodities have relative quality assurance.</td>
</tr>
<tr>
<td>E. Community security</td>
<td>F1. Novel commodities</td>
<td>Take part in on-line group buying for extraordinary commodities that are unavailable on the market.</td>
</tr>
<tr>
<td></td>
<td>G1. Lower price</td>
<td>Take part in on-line group buying for buying more commodities at lower prices.</td>
</tr>
<tr>
<td></td>
<td>G2. Free gift</td>
<td>Take part in on-line group buying for gifts.</td>
</tr>
<tr>
<td>F. On-line group buying diversity</td>
<td>H1. Convenient receiving</td>
<td>Take part in on-line group buying for convenient receiving and delivery services.</td>
</tr>
<tr>
<td>G. On-line group buying price</td>
<td>H2. Clear information presentation</td>
<td>Take part in on-line group buying for explicit website information.</td>
</tr>
<tr>
<td>H. On-line group buying convenience</td>
<td>I1. Simple shopping process</td>
<td>Take part in on-line group buying for simple shopping process.</td>
</tr>
</tbody>
</table>

Source: (Hsu et al., 2014; Nepomuceno et al., 2014; See-Ts and Ho, 2014; Zhao and Zhu, 2014)

2.3. Dynamic Performance Indicator Sequencing - Integrate Grey Correlation Sorting with Dynamic VIKOR

177
Considering the prosperity countermeasure signal, the performance indicator influencing the uncertainty on-line group buying community is analyzed in Table 1. The prosperity countermeasure signal uses five signal lights to represent a situation of prosperity. The cross-check scores are shown in the brackets following the light signal (Ou-Yang and Chuang, 2007) (1) red light (38-45 points) represents overheating; (2) yellow-red light (32-37 points) represents activated prosperity; (3) green light (23-31 points) represents stable prosperity; (4) yellow-blue light (17-22 points) represents poor prosperity; (5) blue light (9-16 points) represents business recession (Deng, 1982; Golmohammadi and Mellat-Parast, 2012; Zhu and Hipel, 2012; Chien, 2015).

The evaluation process of this project follows the dynamic VIKOR multicriterion evaluation method, where \( w_j \) considers the prosperity countermeasure signal, and performance is evaluated according to the weights of various phases and sequencing processes (Hung et al., 2013).

Step 1: Evaluate the scores of ratees in various periods
If each reviewer uses a triangular intuitionistic fuzzy number to represent the result of the criteria, according to the selection criteria of ratees, the denotation is \( (a,b,c;x,y) \), where \( a,b,c \) represent the fuzzy value of a single criterion result of a ratee, the triangular fuzzy number display is flexible, \( x,y \) represent the degree of consent and objection of the decision maker, respectively, and the best result of the assessment score (ideal solution) is represented by \((10,10,10,1,0)\).

Step 2: Calculate the distance between each ratee's criteria results and the best score
According to Szmidt and Kacprzyk (2000) when the distance between the criteria \( D_i (i = 1,2,...,n) \) result in each phase \( T_i (k = 1,2,...,p) \) of each ratee \( D_j (j = 1,2,...,m) \) and the optimal value is calculated, the distance matrix \( \delta_{ki} (k = 1,2,...,p) \) of various phases can be obtained.

Step 3: Evaluation by the VIKOR multicriterion evaluation method
(1) Determine the best value and worst value in the criterion function
Each ratee \( B_j (j = 1,2,...,m) \) is evaluated under evaluation criteria \( D_i (i = 1,2,...,n) \) under No. \( D_i \) evaluation criterion, the performance evaluation value of alternative \( B_j \) is represented by \( R_{ij} \), the best value \( R^+_i \) and worst value \( R^-_i \) of criterion function vary with the effectiveness (larger-the-better) criterion and cost (smaller-the-better) criterion, expressed as Eqs. (15)–(16).

If there are \( m \) alternatives \( B = [B_j]_{j=1,...,m} \), the evaluation is implemented under \( n \) evaluation criteria \( D = [D_i]_{i=1,...,n} \), under No. \( i \) evaluation criterion, the performance evaluation value of alternative \( B_j \) is represented by \( R_{ij} \). If the criterion is the effectiveness (larger-the-better) criterion, the best value \( R^+_i \) and worst value \( R^-_i \) are expressed as Eq. (15).

\[
R^+_i = M \max_j R_{ij}, R^-_i = \min_j R_{ij}
\]  

(15)

On the contrary, if the criterion is the cost (smaller-the-better) criterion, the best value \( R^+_i \) and worst value \( R^-_i \) are expressed as Eq. (16).

\[
R^+_i = \min_j R_{ij}, R^-_i = \max_j R_{ij}
\]  

(16)

(2) Normalization
When the distance between the triangular intuitionistic fuzzy result value of each ratee under the selection criteria, and the best result \((10,10,10,1,0)\) is obtained, the normalized value of ratees under criteria is calculated by Eq. (17).

The normalized value of \( B_j \) scheme under \( D_i \) criterion is represented by \( V_i(B_j) \), expressed as Eq. (17):

\[
V_i(B_j) = \frac{R^+_i - R^-_i}{R^+_i - R^-_i}
\]  

(17)

(3) Calculate \( A_j \) and \( C_j \) values
The criterion normalized value of a ratee in each phase is multiplied by the weight of the criterion, according to Eq. (18). Finally, the weighted normalized values are summed up to obtain \( A_j \), and the criterion normalized weights of ratees are compared according to Eq. (18) where the maximum value is \( C_j \), expressed as Eq. (19).

\[
A_j = \sum_i w_i \frac{R^+_i - R^-_i}{R^+_i - R^-_i}, \forall j
\]  

(18)

\[
C_j = \max_i \left[ \sum_i w_i \frac{R^+_i - R^-_i}{R^+_i - R^-_i} \right], \forall j
\]  

(19)

(4) Select \( A^+, A^-, C^+, C^- \) values
When \( A_i \) and \( C_i \) are obtained, \( A^*, A^* \) are selected from \( A_i \) of criteria, \( A^* \) represents the minimum \( A_i \) value \( \left( A^* = \min A_i \right) \). \( A^* \) represents the maximum \( A_i \) value \( \left( A^* = \max A_i \right) \). \( C^*, C^* \) are selected from \( C_i \) of criteria, \( C^* \) represents the minimum \( C_i \) value \( \left( C^* = \min C_i \right) \), \( C^* \) represents the maximum \( C_i \) value \( \left( C^* = \max C_i \right) \).

(5) Calculate \( U_j \) value

The \( U_j \) value of each scheme is calculated according to Eq. (20). The decision mechanism coefficient \( \alpha \) (weight) is set as 0.5, in order to simultaneously maximize group utility and minimize specific regret.

\[
U_j = \alpha \left[ \frac{\left( A_i - A^* \right)}{\left( A^* - A^* \right)} \right] + (1 - \alpha) \left[ \frac{\left( C_i - C^* \right)}{\left( C^* - C^* \right)} \right] \quad \forall j \quad (20)
\]

Step 4: Implement weighting and sequencing according to the weights of various periods

Chen and Li (2011) proposed the multiple attribute decision making problem, with \( m \) decision making sides \( B_j (j = 1, 2, ..., m) \) and \( n \) evaluation attributes \( D_i (i = 1, 2, ..., n) \), where the weight \( w_j (1, 2, ..., n) \) of each attribute should be considered, and the probably of different performances in multiple phases \( T_k \) must be considered, in order to perfect the evaluation results of various proposals. In this project, as the \( U_j \) value is smaller-the-better criterion (smaller is better), polarity inversion is implemented for the weight of period, and the converted actual weight is represented by \( w_{jk}^* (k = 1, 2, ..., p) \), expressed as Eq. (21):

\[
w_{jk}^* = \frac{1 + w_j}{\sum_{k=1}^p (1 + w_j)} \quad (21)
\]

The \( U_j \) value of various phases is represented by \( U_{jk} (k = 1, 2, ..., p) \), and multiplied by actual weight \( W_{jk}^* (k = 1, 2, ..., p) \), in order to obtain the actual weight \( U_{jk}^* \) of each period, expressed as Eq. (22).

\[
U_{jk}^* = w_{jk}^* \times U_{jk} \quad (22)
\]

The weighted \( U_{jk}^* \) values of the phases are summed up to obtain the \( T \) value, expressed as Eq. (23):

\[
T = \sum_{k=1}^p U_{jk}^* \quad (23)
\]

When the smaller weighted sum \( T \) of the \( U_{jk}^* \) value of various periods is better, the total sequence of the full phase can be obtained.

This project is based on dynamic VIKOR, with the prosperity countermeasure signal and analysis results as shown in Table 1. The degree of relationship among sub-systems or elements could be evaluated through grey relational analysis, and the important influential factors of the development trend are determined in order to learn the major features of the system through the following steps, as shown in Eq. (24)-(29) (Deng, 1982; Chien, 2015).

Step 1: Normalize original data. Normalize by dividing the original data \( x_i (k) \) with the mean value of the sequence shown in Eq. (1):

\[
r_i(k) = \frac{x_i(k)}{\sum_{k=1}^N x_i(k)} \quad i = a, ..., d \quad k = A, ..., N \quad (24)
\]

Step 2: Designate the standard sequence and calculate the difference sequence. Take the mean value as a standard sequence, i.e. sequence 0; the difference sequence \( \Delta_{ki} (k) \) indicates the absolute difference of elements \( k \) between the other sequence \( i \) and the standard sequence 0, as expressed in Eq. (5):

\[
\Delta_{ki}(k) = |r_i(k) - r(k)| \quad i = 1, 2, 3, ... \quad k = A, ..., N \quad (25)
\]

Step 3: Calculate maximal difference \( \Delta_{\max} \) and minimal difference \( \Delta_{\min} \), as expressed in Eqs. (6) and (7).

\[
\Delta_{\max} = \max_{r \neq k} \Delta_{ki}(k) \quad (26)
\]

\[
\Delta_{\min} = \min_{r \neq k} \Delta_{ki}(k) \quad (27)
\]

Step 4: Calculate grey relational coefficient: \( \gamma_{0i}(k) \). The relational coefficient: \( \gamma_{0i}(k) \) is defined below, of which \( \zeta \) is the adjustment factor, as shown in Eq. (8).

\[
\gamma_{0i}(k) = \frac{\Delta_{\min} + \zeta \cdot \Delta_{\max}}{\Delta_{0i}(k) + \zeta \cdot \Delta_{\max}} \quad (28)
\]

Step 5: Calculate the grey relationship \( \Gamma_{0i} \) between each sequence and the standard sequence. The grey relationship \( \Gamma_{0i} \) is defined as in Eq. (9).
\[ \Gamma_{\alpha} = \sum_{k=1}^{N} \frac{\gamma_{\alpha}(k)}{N} \]  

(29)

Step 6: Conduct sequencing according to the grey relationship.

3. Results Analysis

According to the 863 data collected from group-buying websites, as shown in Table 2 the military, civil servants, and teachers account for a high proportion of group buying. The fitness values of various populations, as calculated by Eq. (11) are greater than 95%, meaning the extracted term is well effective at classifying the document as the original type. The performance indicators of different populations are calculated using the dynamic VIKOR method. Afterwards, the key performance indicator of each population is calculated by grey correlation sorting. Students and homemakers care most about minimum price; the freelance and service industry care most about convenient receiving; the service industry cares most about simple shopping processes; the financial services/insurance industry cares most about financial risk; the military, civil servants, and teachers care most about reliable group-buying websites; the information industry cares most about the master buyer's reputation and evaluation. The analysis results fully show the life contexts of different populations, as well as the on-line group buying content they are most interested in.

<table>
<thead>
<tr>
<th>Population type</th>
<th>Proportion</th>
<th>Fitness</th>
<th>( T ) (value, indicator)</th>
<th>Performance (Value, Key indicator)</th>
<th>( \Gamma_{\alpha} ) (Value, key indicator)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students</td>
<td>4.53%</td>
<td>96.11%</td>
<td>0.26 G1,G2,F1</td>
<td>0.56 G1</td>
<td></td>
</tr>
<tr>
<td>Homemakers</td>
<td>13.26%</td>
<td>97.18%</td>
<td>0.13 G1,E2,C2</td>
<td>0.58 G1</td>
<td></td>
</tr>
<tr>
<td>Freelance</td>
<td>12.25%</td>
<td>96.21%</td>
<td>0.18 A2,H1,E2</td>
<td>0.66 H1</td>
<td></td>
</tr>
<tr>
<td>Manufacturing industry</td>
<td>6.26%</td>
<td>95.13%</td>
<td>0.11 E2,H1,J2</td>
<td>0.71 H1</td>
<td></td>
</tr>
<tr>
<td>Service industry</td>
<td>16.38%</td>
<td>98.32%</td>
<td>0.28 E2,I2,I1</td>
<td>0.68 I2</td>
<td></td>
</tr>
<tr>
<td>Financial services / insurance industry</td>
<td>8.25%</td>
<td>95.29%</td>
<td>0.19 G2,D2,D1</td>
<td>0.79 D1</td>
<td></td>
</tr>
<tr>
<td>Military, civil servants and teachers</td>
<td>28.69%</td>
<td>97.28%</td>
<td>0.09 E2,D2,D1</td>
<td>0.87 E2</td>
<td></td>
</tr>
<tr>
<td>Information industry</td>
<td>10.38%</td>
<td>95.16%</td>
<td>0.27 D2,D1,C1</td>
<td>0.81 C1</td>
<td></td>
</tr>
</tbody>
</table>

Source: This research.

4. Conclusion

The life contexts and habits of different populations of on-line group buying match the results of this study; student and homemakers care most about minimum price, because the economic situation of this population is relatively controlled; the freelance and service industry care most about convenient receiving, because their work places are inconvenient for receiving; the service industry cares most about simple shopping processes, because they have less spare time; the financial services/insurance industry cares most about financial risk, due to their profession; the military, civil servants, and teachers care most about reliable group-buying websites, due to their cautious behavior style; the information industry cares most about master buyer's reputation and evaluation, due to sensitivity to information.

References


