A Fuzzy Decision Support Method of Product Purchase Based on Text Reviews

Yifan Yang [a], Pu Ji [b*]
[a] 202121702002@stu.hebut.edu.cn, [b*] 2019086@hebut.edu.cn

School of Economics and Management, Hebei University of Technology, Tianjin, China

Abstract: Electronic commerce platforms provide decision support for their users due to information overload. How to rank products through online reviews is an important research topic. And, hesitancy and fuzziness exist in text reviews. This paper applies fuzzy logic and methods to solve decision support problems. Given this, the proposed method introduces probabilistic intuitionistic hesitant fuzzy numbers (PIHFNs) to characterize text reviews. A PIHFN can reflect multiple positive, neutral and negative information in text reviews. In addition, this paper proposes a distance formula based on PIHFNs, which is used in the proposed method to provide decision support. The proposed method generates suggested products through the technique for order preference by similarity to an ideal solution. The performance of the method in terms of accuracy is measured by the total relative difference metric. The comparisons and experiments are further conducted to illustrate the advantages of the proposed method. Converting the identified positive, neutral, and negative sentiment orientations into PIHFNs is a new idea for processing online reviews. The proposed method can be applied to tackle the decision support problem accurately.

Keywords: decision support; text reviews; probabilistic intuitionistic hesitant fuzzy numbers; electronic commerce

1 INTRODUCTION

With the development of electronic commerce, more and more people choose to shop online and leave their reviews. Research indicates that reviews of users generally have a great impact on the purchase decisions of other individuals. However, it is difficult for people to make purchase decisions because of huge reviews.

Online reviews can be used in decision support. Especially text reviews. Extant research study decision support is based on text reviews by Sentiwordnet to attach sentiment scores to words and then calculate sentiment scores of the sentences so that text data is transformed into structural data [1].

Fuzziness and hesitancy exist in text reviews. For example, a text review about a restaurant is as follows: fruit juice drinks delicious, the chicken tastes good, the fish is delicious, but the beef is so tough. Considering this, Liu et al. [2] utilized the intuitionistic fuzzy numbers to characterize fuzziness in reviews. Zhang et al. [3] used the hesitation fuzzy set to describe fuzziness and hesitancy in online reviews. However, the hesitant fuzzy number can not reflect the probability of the occurrence of “delicious” in the above text reviews. PIHFNs, which are composed of intuitionistic fuzzy numbers and probabilities, can be introduced to reflect

ICICA 2022, December 02-04, Chongqing, People's Republic of China

DOI 10.4108/eai.2-12-2022.2327982

information in the above reviews. Therefore, to support the consumers' purchase decisions, it is necessary to propose an algorithm for identifying sentiment orientations in reviews and ranking the products based on sentiment analysis technology and PIHFNs theory.

In this paper, we assign positive and negative scores to words according to Sentiwordnet and then extract and classify sentiment words. After that PIHFNs are firstly utilized to depict multiple positive, neutral, and negative information in text reviews. Furthermore, we define the distance measure of PIHFNs. This measure will be used in the technique for order preference by similarity to an ideal solution (TOPSIS). TOPSIS is introduced to rank products and generate decision supports. This study verifies the performance of the proposed method with data from Amazon.

The rest of this paper is organized as follows. Section II briefly reviews the basic relevant theories. In Section III, we define the distance measure of PIHFSs. Section IV presents our proposed method. Two numerical examples are given in Section V. Finally, we conclude in Section VI.

2 RELATED WORK

This section introduces the notions of some fuzzy sets which are related to PIHFSs. The definition of IFSs is introduced here:

Definition 1[4]: Let X be a fixed set, then an intuitionistic fuzzy set (IFS) associated with X can be denoted as:

$$A = \{\langle x, \mu_A(x), \gamma_A(x)\rangle | x \in X\} \tag{1}$$

where the mappings $\mu_A(\cdot)$ and $\gamma_A(\cdot): X \to [0,1]$ are the membership and non-membership function of x in A respectively, and satisfy $0 \le \mu_A(x) + \gamma_A(x) \le 1, \forall x \in X$. Furthermore, $\pi_A(x) = 1 - \mu_A(x) - \gamma_A(x)$ is used to denote indeterminacy degree (or hesitant degree), which can be considered as an intuitionistic index of x in X , and in the meantime $\pi_A(x)$ satisfies $0 \le \pi_A(x) \le 1$. One element of an IFS is called an intuitionistic fuzzy number (IFN). For convenience, an IFN is denoted by $\langle \mu_A, \gamma_A\rangle$.

Hesitancy exists in human cognition. One decision-maker may hesitate among multiple IFNs at one time. IHFSs can be used to express such decision information. The definition of IHFSs is presented as follows:

Definition 2[5]: Let $P^2([0,1])$ be the family of subsets of some different two tuples in $[0,1]^2$, a hesitant intuitionistic fuzzy set (HIFS) A on X can be obtained as

$$A = \{\langle x, h_A(x)\rangle \mid x \in X\} \tag{2}$$

where the mapping

$$h_A(\cdot) : U \to P^2([0,1])$$

$$x \to \left\{(\mu_A^i(x), \gamma_A^i(x)) : 1 \le i \le n_A(x)\right\}$$

is called the hesitancy function and represents all possible two tuples with the first argument are membership degree and the second variable are non-membership degree. It also satisfies $0 \le$

$\mu_A^i(x), \gamma_A^i(x) \leq 1$ and $0 \leq \mu_A^i(x) + \gamma_A^i(x) \leq 1$, $\forall 1 \leq i \leq n_A(x)$ for all $x \in X$. And hesitant indeterminacy function can be given as $\pi_A(x) = \left\{\pi_A^i(x) = 1 - \mu_A^i(x) - \gamma_A^i(x) : 1 \leq i \leq n_A(x)\right\}$, which reflects the whole intuitionism of the object x in A. Furthermore, $\pi_A(x)$ satisfies $0 \leq \pi_A^i(x) \leq 1,\ 1 \leq i \leq n_A(x)$. For the sake of simplicity, all HIFSs on X are represented as HIFS(X) and the set $\alpha = \{(\mu_\alpha^i, \gamma_\alpha^i) : 1 \leq i \leq n_\alpha\}$ is called the hesitant intuitionistic fuzzy number (HIFN).

IFNs in a HIFN may have different possibilities, the PIHFNs reflect the possibility of each IFN. The definition of PIFHSs is introduced here:

Definition 3: Let X be a fixed set, then a probabilistic intuitionistic hesitant fuzzy set (PIHFS) associated with X can be denoted as:

$$I = \{\langle x, h_A(x), p_A(x)\rangle | x \in X\} \tag{3}$$

where $h_A(x) = \langle \mu_A(x), \gamma_A(x)\rangle$ is an IFN, $p_A(x)$ denotes the occurrence probability of $h_A(x)$, $0 \leq p_A(x) \leq 1$ with $\sum_{i=1}^{\#h} p_A(x) \leq 1$ and $\#h$ is the number of elements in I. One element of I, denoted by $\langle h_A(x), p_A(x)\rangle$, is a probabilistic intuitionistic hesitant fuzzy number (PIHFN).

3 THE DISTANCE BETWEEN PIHFNS

The definition of distance between PIHFNs.

Definition 4: For two PIHFNs$H_1 = \{\langle \mu_1^l, \gamma_1^l, p_1^l\rangle \mid l = 1,2,\cdots,L\}$ and $H_2 = \{\langle \mu_2^l, \gamma_2^l, p_2^l\rangle \mid l = 1,2,\cdots,L\}$ defined on X, we define the distance between H_1 and H_2 as:

$$d(H_1, H_2) = \frac{\sum_{l=1}^{L} |\mu_1^l - \mu_2^l|^2 p_1^l p_2^l + \sum_{l=1}^{L} |\gamma_1^l - \gamma_2^l|^2 p_1^l p_2^l}{2L} \tag{4}$$

where L represents the length of H_1 or H_2.

The distance measure proposed above satisfies the axiomatic statement given below:

Theorem 1 Let $H_1 = \{\langle \mu_1^l, \gamma_1^l, p_1^l\rangle \mid l = 1,2,\cdots,L\}$ and $H_2 = \{\langle \mu_2^l, \gamma_2^l, p_2^l\rangle \mid l = 1,2,\cdots,L\}$ be two PIHFNs, then the distance measure satisfies the following conditions:

(P1) $0 \leq d(H_1, H_2) \leq 1$

(P2) $d(H_1, H_2) = d(H_2, H_1)$

(P3) $d(H_1, H_2) = 0$ *if and only if* $H_1 = H_2$

4 THE DECISION SUPPORT METHOD

4.1 Characterize Text Reviews

To express the word accurately, we apply PHIFSs to depict text reviews of products.

Firstly, we represent products' text reviews by HIFSs. For product x_i, let

$$H(x_i) = \{\{(\mu_i^1, \gamma_i^1), (\mu_i^2, \gamma_i^2), \cdots, (\mu_i^n, \gamma_i^n)\}|$$

$$x_i \in X, i = 1,2,\dots,I\} \quad (5)$$

where(μ_i^j, γ_i^j)indicates a positive score and a negative score respectively on the j^{th}sentiment word and n refers to the number of sentiment words in reviews of product x_i. The positive score and negative score with the most commonly used meanings are obtained from Sentiwordnet. In addition, when these scores of a word are not 0, it is considered a sentiment word. Then, we carry out a series of simplified operations appropriately to reduce calculation. According to the possible membership degree μ and non-membership degree γ and hesitant degree π, the HIFS of a product is divided into three HIFSs. These express respectively positive sentiment words set, negative sentiment words set and neutral sentiment words set. If $\mu > max\{\gamma, \pi\}$ for a word, this word is assigned to Hesitant Intuitionistic Fuzzy Positive Number (HIFPN); and if $\gamma > max\{\mu, \pi\}$, this word is assigned to Hesitant Intuitionistic Fuzzy Negative Number (HIFNN); and then if $\pi > max\{\mu, \gamma\}$, this word is assigned to Hesitant Intuitionistic Fuzzy Neutral Number (HIFON). Because HIFSs can be used to aggregate each sentiment word value to get a comprehensive expression value. We apply the aggregation operation developed by Xu [6] to HIFPN, HIFNN, and HIFON respectively. Aggregation operation is given as:

Given that $H(x_i) = \{(\mu_i^1, \gamma_i^1), (\mu_i^2, \gamma_i^2), \cdots, (\mu_i^n, \gamma_i^n)\}|$

$x_i \in X, i = 1,2,\dots,I\}$, let $c^1, c^2, \cdots, c^n$ be a set of IFNs with $c^j = (\mu^j, \gamma^j)$, and w^j be the weight of c^j with $w^j \geq 0$ for $j = 1,2,\cdots,n$ and $\sum_{j=1}^{n} w^j = 1$. Then, intuitionistic fuzzy arithmetic averaging (IFWA) operation expresses as:

$$\text{IFWA}(c^1, \cdots, c^n) = \langle 1 - \prod_{j=1}^{n}(1-\mu^j)^{w^j}, \prod_{j=1}^{n}(\gamma^j)^{w^j} \rangle \quad (6)$$

where due to the impact of each sentiment word is considered to be the same, let $w^j = \frac{1}{n}$ for $j = 1,2,\cdots,n$. After that HIFPN, HIFNN, and HIFON of each product are represented by IFPN, IFNN, and IFON. And let probabilities of IFPN, IFNN, and IFON as follows:

$$p_{\text{IFPN}} = \frac{N_{pos}}{N} \quad (7)$$

$$p_{\text{IFNN}} = \frac{N_{neg}}{N} \quad (8)$$

$$p_{\text{IFON}} = \frac{N_{obj}}{N} \quad (9)$$

where N is the number of sentiment words in reviews of a product. N_{pos}, N_{neg}, and N_{obj} represent the length of HIFPN, HIFNN, and HIFON of a product respectively.

Next, we utilize those IFNs and probabilities to constitute PHIFSs expressing text reviews of products. Then,

$$H(p(x_i)) = \{(\mu_i^1, \gamma_i^1, p_i^1), (\mu_i^2, \gamma_i^2, p_i^2), (\mu_i^3, \gamma_i^3, p_i^3)\}|$$
$$x_i \in X, i = 1,2,\dots,I\} \quad (10)$$

where $(\mu_i^1, \gamma_i^1, p_i^1)$, $(\mu_i^2, \gamma_i^2, p_i^2)$, and $(\mu_i^3, \gamma_i^3, p_i^3)$donate the results of positive, negative and neutral words aggregation and relative probabilities. Hence, $H(p(x_i))$ is employed to donate the text reviews of product x_i.

4.2 Generate Suggested Products

To generate suggested products, we utilize the extended TOPSIS method to deal with hesitant and intuitionistic fuzzy information and employ the distance measure of PIHFSs to gain the final ranking of the products. TOPSIS, a kind of method to solve decision problems, has been extended to deal with fuzzy information in the last few years [7]. The idea of this method is to choose the alternative with the shortest distance from the positive ideal solution(PIS) and the farthest distance from the negative ideal solution(NIS). However, most existing research studies on TOPSIS methods based on fuzzy sets, hesitant fuzzy sets, and intuitionistic fuzzy sets[8]. The TOPSIS method based on probabilistic intuitionistic hesitant fuzzy sets has not been proposed.

The PIS of PIHFSs, denoted by A^+, and the NIS of PIHFSs, denoted by A^-. In this paper, that can be defined as follows:

$$A^+ = (1,0,1),(1,0,1),(1,0,1)$$

$$A^- = (0,1,1),(0,1,1),(0,1,1)$$

The distances $d_i^+$and d_i^- of product x_i from $A^+$and A^-, respectively, are derived from

$$d_i^+(H\,(p(x_i)),A^+) = \frac{\sum_{l=1}^{L}|\mu_1^l-\mu_2^l|^2 p_1^l p_2^l+\sum_{l=1}^{3}|\gamma_1^l-\gamma_2^l|^2 p_1^l p_2^l}{2L} \quad (11)$$

$$d_i^-(H\,(p(x_i)),A^-) = \frac{\sum_{l=1}^{L}|\mu_1^l-\mu_2^l|^2 p_1^l p_2^l+\sum_{l=1}^{3}|\gamma_1^l-\gamma_2^l|^2 p_1^l p_2^l}{2L} \quad (12)$$

where $L = 3$ and $i = 1,2,\cdots,I$

The relative closeness coefficient of a product x_i concerning PIS of PIHFSs A^+ is defined as follows:

$$C_i = \frac{d(H\,(p(x_i)),A^-)}{d(H\,(p(x_i)),A^-)+d(H\,(p(x_i)),A^+)} \quad (13)$$

where $0 \le C_i \le 1, i = 1,2,\cdots,I$. Obviously, if the product x_i is closer to the value of A^+ and farther from the value of A^- as C_i is close to 1. Therefore, we utilize the closeness coefficient C_i to determine the ranking order of all products.

4.3 Steps of the Decision support Method

Decision support method is mainly divided into the following modules and Fig.1 shows the steps of the proposed method:

(1) Acquiring reviews data of products, save data to the database.

(2) Sentiment words extraction. According to the positive and negative scores of words provided by Sentiwordnet, sentiment words are extracted from text reviews.

(3) HIFSs construction. Sentiment words are divided into three sets and respectively used to construct HIFSs (HIFPN, HIFNN, and HIFON) expressing positive sentiment words set, negative sentiment words set and neutral sentiment words set respectively by comparing sentiment scores.

(4) HIFSs aggregation. Aggregating these HIFSs and computing relevant probabilities, then converting them to a PHIFS $H\,(p(x_i))$ to describe a product x_i.

(5) Distance between PHIFS of a product and PIS and NIS calculation by formula (11) and (12).

(6) Closeness coefficient calculation by formula (13). Finally, we utilize the closeness coefficient C to determine the ranking order of all products.

5 ILLUSTRATIVE EXAMPLE

In this section, we present different experiments to evaluate the performance of our proposed method.

5.1 Experimental Design

In this paper, the proposed method has been developed by the python and the Anaconda3 platform. The main part of the data set includes the reviews, hours, and ratings of 18 products, distributed in 2 categories from Amazon. This data set includes approximately 15,000 texts. Besides, the process of marking sentiment polarity scores of words has been performed using Sentiwordnet. For example, the most commonly used meaning of "nice" has a positive score of 0.875 and a negative score of 0.125, so the data form in a HIFS is (0.875, 0.125). To verify the effectiveness of our proposed method, we compare it with the method of using IFSs to represent reviews [2] and the method of average star rating. In addition, we calculate the average star rating of each product by obtaining the star rating in each review. To this end, sentiment words were extracted from reviews of each product and categorized into positive, negative, and neutral word sets as shown in Table 1. Besides, HIFPN, HIFNN, and HIFON were utilized to represent these respectively then converted to PHIFSs by aggregation operations and probability calculations as shown in Table 2. Next, we obtained the distances $d_i^+(H\,(p(x_i)), A^+)$ and $d_i^-(H\,(p(x_i)), A^-)$ as shown in Table 2 between PHIFSs and A^+ and A^- by distance formulas (11) and (12) .Finally, we compute the closeness coefficient of each product by formula (13) to rank.

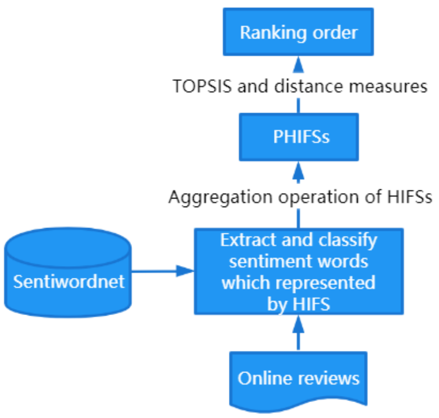

Figure 1 The general framework of a fuzzy decision support method of product purchase based on text reviews

Table 1 The set of positive, negative, and neutral words of a product

Word set	Sentiment word
Positive words	love, good, easy, perfect, …,super
Negative words	deceiving, wrong, terrible, awful, ⋯,warning
Neutral words	soft, exactly, pretty, dry, …,pink

Table 2 The value of HIFPN, HIFON and HIFNN of a product

HIFSs		PHIFSs
HIFPN	{(0.625,0.375) ⋯ (0.625,0.125),(0.500,0.125)}	(1.000,0.000,0.318)
HIFON	{(0.125,0.000) ⋯ (0.125,0.375),(0.250,0.250)}	(0.051,0.604,0.095)
HIFNN	{(0.000,0.625) ⋯ (0.125,0.500),(0.000,0.750)}	(0.132,0.000,0.586)

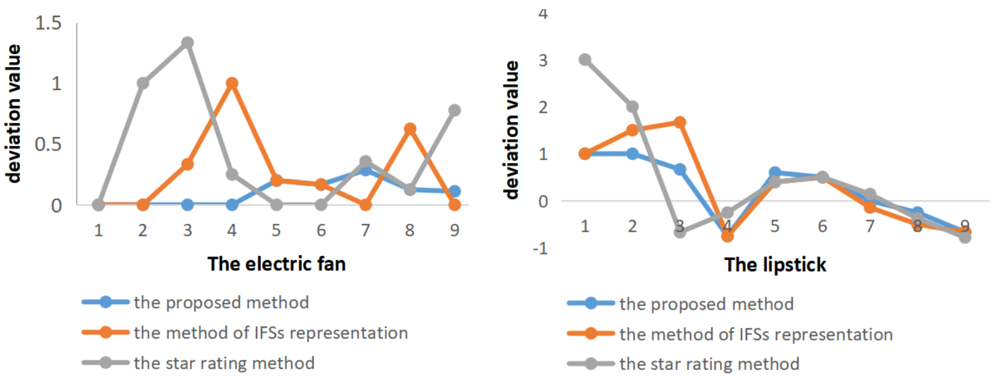

Figure 2 Relative differences between the ranking result of each approach and the baseline order of the lipstick and the electric fan

Table 3 The total deviations of the three approaches in each product

Product	The proposed approach	The method of IFSs representation	The star rating method
The lipstick	5.433	7.126	8.112
The electric fan	0.889	2.325	3.843

5.2 Experimental Results

To accurately reflect the efficacy of this method, the relative differences between the ranking results of the three experiments and the baseline order were determined. The baseline order respectively represents the sales order of each product on Amazon about the lipstick and the electric fan. And, the order of products was the result of ranking according to the baseline order. The relative difference was computed as [9]:

$$d_i = \frac{r_i - B_i}{B_i} \quad (14)$$

where r_i represents the ranking order of the experiment of product x_i under a product type, and B_i represents the baseline order. The relative differences of each product were given in Fig.2. For determining which approach exhibited the best performance, the total difference between the ranking result of each approach and the baseline order was calculated as:

$$SumDiv = \sum_{i=1}^{n} \left| \frac{r_i - B_i}{B_i} \right| \quad (15)$$

where $SumDiv$ denotes the sum of all of the absolute values of the relative differences or the total difference of the approach. The total differences between the three approaches to the lipstick and the electric fan were listed in Table 3. The proposed approach exhibited higher performance than others' methods. Therefore, PIHFN is utilized to depict multiple positive, neutral and negative information in text reviews that can generate accurate results.

6 Conclusions

In this paper, due to the hesitancy and fuzziness in text reviews, we propose a new decision support method. PIHFNs are firstly introduced to characterize text reviews. Moreover, we define the distance measure of PIHFNs. The distance measure is applied in the decision support method. Furthermore, the TOPSIS method based on PIHFNs is proposed to rank products and provide decision support. Finally, we verify the performance of the proposed method with reviews from Amazon. We compare our method with the method of using IFSs to represent reviews and the method of average star rating, then the performance of methods in terms of accuracy is measured by the total relative difference metric. Results show that the proposed method performs well in effectiveness. Therefore, it is important for developing and enriching theories and methods for ranking products through online reviews. In future work, the issue of how to further improve the accuracy of ranking through reviews and auxiliary information and the situation that new products with few online reviews need to be considered.

Acknowledgment. This work is supported by National Natural Science Foundation of China (No. 72001068).

REFERENCES

[1] Z. Abbasi-Moud, H. Vahdat-Nejad, and J. Sadri, "Tourism recommendation system based on semantic clustering and sentiment analysis," *Expert Systems with Applications,* vol. 167, 2021, doi: 10.1016/j.eswa.2020.114324.

[2] Y. Liu, J.-W. Bi, and Z.-P. Fan, "Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory," *Information Fusion,* vol. 36, pp. 149-161, 2017.

[3] D. Zhang, C. Wu, and J. Liu, "Ranking products with online reviews: A novel method based on hesitant fuzzy set and sentiment word framework," *Journal of the Operational Research Society,* vol. 71, no. 3, pp. 528-542, 2019, doi: 10.1080/01605682.2018.1557021.

[4] K. T. Atanassov, "Intuitionistic Fuzzy Sets," *Fuzzy Sets and Systems,* vol. 20, pp. 87–96, 1986.

[5] D. Yao and C. Wang, "Hesitant intuitionistic fuzzy entropy/cross-entropy and their applications," *Soft Computing,* vol. 22, no. 9, pp. 2809-2824, 2017, doi: 10.1007/s00500-017-2753-x.
[6] Z. Xu and R. R. Yager, "Some geometric aggregation operators based on intuitionistic fuzzy sets," *International Journal of General Systems,* vol. 35, no. 4, pp. 417-433, 2006, doi: 10.1080/03081070600574353.
[7] C.-C. Hung and L.-H. Chen, "A Fuzzy TOPSIS Decision Making Model with Entropy Weight under Intuitionistic Fuzzy Environment," in *Proceedings of the International Multi-Conference of Engineers and Computer Scientists IMECS*, Hong Kong, 2019.
[8] K. Palczewski and W. Sałabun, "The fuzzy TOPSIS applications in the last decade," *Procedia Computer Science,* vol. 159, pp. 2294–2303, 2019.
[9] H.-y. Zhang, P. Ji, J.-q. Wang, and X.-h. Chen, "A Neutrosophic Normal Cloud and Its Application in Decision-Making," *Cognitive Computation,* vol. 8, no. 4, pp. 649-669, 2016, doi: 10.1007/s12559-016-9394-8.