

Aggregation Operators in Hesitant Fuzzy Set for Decision Making



Roopa Chandrika R, Gowri Ganesh N.S, Mummoorthy A, Gayathri M

Abstract: Uncertainty is prevalent in a wide range of real-world issues. The fuzzy sets, vague sets or intuitionistic fuzzy sets are widely used in recent years for decision making and various analysis where uncertainty is predominant. An extension of fuzzy sets is Hesitant Fuzzy Sets, which deals with ambiguous situations that arise when determining an element's membership degree in a set. Researchers have defined various ideas, extensions, aggregation operators, and measurements to deal with reluctant information as a result of this new approach. Machine learning algorithms are also exploiting hesitant fuzzy sets for better decision making.

Keywords: Aggregation Operators, Hesitant Reviews, Hesitant Fuzzy Sets, Sentiment Analysis

I. INTRODUCTION

Cantor introduced the concept of set as a collection of elements and each element having a particular property that identifies its individuality in the set. The value of an element belonging to a set uses a bivalent concept such that the element having value one belongs to the set and otherwise non member of the set. Later Zadeh introduced the concept of fuzzy sets assigning degree of membership to each element belonging to the set that can have a value between 0 and 1. Similar to the set theory, an element with the value 1 is completely belonging to the set and 0 otherwise. The elements with the membership value between 0 and 1 will be partially belonging to a set. In later studies, Atanassov introduced the idea of vague sets or intuitionistic fuzzy set where the sum of degree of all elements in the set will be equal to 1 or less than 1 [1]. The fuzzy sets, vague sets or intuitionistic fuzzy sets are widely used in recent years for decision making and various analysis where uncertainty is predominant. Uncertainty is of two kinds, one is uncertain whether an element is belonging to a set or not and another aspect is the amount of information contained. In order to resolve this, researchers have formulated entropy measures to

quantify the uncertainty. There are challenges yet in decision making tasks because of multiple ideas, multiple criteria from multiple experts. Several decision making models are designed and experimented and one of the significant task in such is the representation of information in uncertain domain. Zadeh proposed fuzzy linguistic model [2] where the linguistic variables are represented in words and richer expressions using natural language instead values like low, moderate, high, very high etc. Models can only express the value of a linguistic variable using a single and simple linguistic phrase, and so cannot convey complex information. But expert judgements are frequently marred by uncertainties or hesitations. Torra [3] introduced the concept of Hesitant Fuzzy Sets (HFS) in which each element was allowed to have multiple values available in the reference/original set.

II. PRELIMINARIES HESITANT FUZZY SETS

A. Definition

Torra [3] has defined a hesitant fuzzy set as , considering the reference set χ and a function (h) is defined on the reference set returns a subset of values 0 or 1. Decisions were modeled using hesitant fuzzy sets in which experts choose an alternative value for an element based on a set of criteria and each criteria was associated with a degree of satisfaction. Experts may struggle to choose between many values when evaluating alternatives or criteria. As a result, aggregation operators are defined in order to aggregate Hesitant Fuzzy Element (HFEs). The decision model has two phases of aggregation and exploitation phase. In the aggregation phase the assessments for each alternative is grouped and in the exploitation phase the best solution is chosen based on the aggregated value. Several aggregation operators like arithmetic mean, weighted fuzzy averaging, geometric mean, generalized fuzzy ordered weighted averaging [4] etc are used by researchers for a better decision process.

B. HFS Aggregation Operators

For a decision problem with many attributes and several experts' opinions, there are generally two ways to aggregate the values. One of them is to aggregate the opinion of the decision makers for each attribute and their alternatives, and then aggregate all the values of the attributes for each alternative. The second method is to aggregate the attribute values given by the decision makers for each alternative, then aggregate the opinion of the decision makers for each alternative. Initially weighted average operator was used and later on versions of fuzzy weighted average were formulated as given below.

- Intuitionistic fuzzy weighted average: For the fuzzy values $(\alpha_1, \alpha_2, \dots, \alpha_n)$ and its associated weights (w_1, w_2, \dots, w_n)

Manuscript received on October 04, 2021.

Revised Manuscript received on October 18, 2021.

Manuscript published on November 30, 2021.

* Correspondence Author

Roopa Chandrika R*, Computer Science, Malla Reddy College of Engineering and Technology, Hyderabad, Telangana, India. Email: mroopachandrika@gmail.com

Gowri Ganesh N.S, Computer Science, Malla Reddy College of Engineering and Technology, Hyderabad, Telangana, India. Email: gowriganesh@gmail.com

Mummoorthy A, Computer Science, Malla Reddy College of Engineering and Technology, Hyderabad, Telangana, India. Email: amummoorthy@gmail.com

Gayathri A, Computer Science, Malla Reddy College of Engineering and Technology, Hyderabad, Telangana, India. Email: gayatricse312@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

and the degree of membership functions as $\mu(x)$ and $\nu(x)$, the weighted average operator is :

$$\left(\prod_{j=1}^n (\mu_{z_j})^{w_j}, 1 - \prod_{j=1}^n (1 - \nu_{z_j})^{w_j} \right) \quad [11]$$

- The geometric mean operator is given as:

$$\left(\prod_{j=1}^n (\mu_{z_j})^{w_j}, 1 - \prod_{j=1}^n (1 - \nu_{z_j})^{w_j} \right) \quad [11]$$

- The hybrid average weight method is given by

$$\left(1 - \prod_{j=1}^n (1 - \mu_{z_{(j)}})^{w_j}, \prod_{j=1}^n (\nu_{z_{(j)}})^{w_j} \right) \quad [11]$$

C. Hfs And Applications

HFSs are being used in decision support systems, machine learning, recommender systems, evaluation where multiple values and multiple criteria needs to be assigned to a parameter. Machine learning uses clustering algorithms for exploratory analysis. It is an unsupervised learning used to model unlabeled dataset. Classification is a considerable aspect in machine learning to deal with real world problems. The classification algorithms categorize objects of similar characteristics or properties into a cluster and those with different properties into another cluster. Some data points or objects are difficult to classify, due to uncertainty or fuzzy environment. Hesitant decisions have to be considered for clustering. Correlation between the objects is considered for classification. N.Chen et.al, have devised correlation coefficient formula [5] and examined the weights of the attributes in HFS for software performance and business risk applications. A correlation matrix is computed and after several compositions an equivalent correlation and λ -cutting matrix are also constructed with λ stating the confidence level. Examining few columns (i,j) in the correlation matrix, if all the elements are the same then the HFSs are of the same type. Applying this principle classification was performed for real world applications as stated above. The clustering algorithm proposed by the authors in the HFS frame has a wide range of applications. Hesitant Fuzzy decision trees (HFDS) method are used by the authors Sahar et.al [6] for a highly imbalanced dataset. In order to balance data sets, K-means clustering algorithm was used to divide at most large class samples into several clusters and labeled each cluster with a new user defined class label. Using HFS, five alternative HFDS were constructed with the goal of performing an unbalanced data classification problem. For expansion of the decision tree the information energy was utilized for defining a novel selecting attribute criterion. Three different strategies were adopted for predicting the class label of the new data and also for aggregating the outcome of five different fuzzy decision trees. Ebrahimpour et al, have analysed a high dimensional dataset that suffered sample sizes that were small and used HFS to resolve the problem that heuristic algorithms could not solve. Maximum Relevance and Minimum Redundancy [6] HFS technique that ensembled ranking algorithms for relevancy measure and ensembled similarity measures to calculate the minimum redundancy measure for feature selection was used.

Digital marketing uses decision support system to understand the sentiments of a customer and to give a reliable recommendation to a product or any reference to a query. Online reviews are generated massively resulting high dimensional data. Tourist evaluations on the internet are the true feelings of visitors after they have travelled, and they have a high reference value for potential tourists making further travel decisions. A potential visitor, on the other hand, will almost certainly be unable to look into a plethora of online evaluations about tourist attractions. People require a recommender system so that he/she can get the most out of his/her trip decision that is appropriate for travelling. Quin et al have proposed TOPIS methodology for analyzing the hesitant reviews and used the principle of deviation minimization [8] to calculate HF-weight entropy information for each criterion. Due to the abundant similar digital items in the market, clients frequently find difficult as a result of the vast selection and, as a result, rely heavily on expert and peer ratings. Fuzzy multi criteria decision making methodology proposed by researchers, used HFS for choosing the best alternative. A novel distance measure [9] based on mean deviation and score function computed by authors seemed to be exhibit more advantageous compared to the traditional methods. The recent research using fuzzy hesitant sets was to find the most significant risk factor in the spread of COVID-19. The researchers have designed a new methodology FAHP and HFS-TOPSIS [10] constructing a comparison matrix associated with fuzzy weights. The linguistic variables are natural language phrases like “not maintaining personal hygiene”, “spread through hospitals and clinic”, “verbal spread”, “not maintaining social distance in public places or public transport”, and “not using quality mask” [10] etc. The authors have used positive and negative ideal solution values and the efficient alternative was chosen such that it is closest to the positive ideal solution.

III. SENTIMENT ANALYSIS USING HFS

Reviews are beneficial to know about the customer's loyalty to the products, services or other entities. Other than business organizations, the public who wish to know or buy a product or to know about the services of any organization take the help of reviews that are already shared by others, for decision making. The reviews of a customer can be well understood with the sentiments that are expressed in their writings or talks. Sentiments, such as opinions, attitudes, views and emotions, are personal experiences of individuals that are not open to impartial observation. They are stated in language that uses subjective opinions which express sentiment analysis. The sentiments of a satisfied customer will be an overwhelmingly comment on the product purchased whereas an unhappy or dissatisfied customer will share their unpleasant experience. In recent years there is an explosion of reviews through social media like Facebook, Instagram, Amazon, Flipkart, online trading websites and Twitter etc. These reviews are expressed as text messages, emoticons and star grading method.



Even though the messages received are voluminous, the contents are short and are expressed in a natural colloquial manner. In such cases mining the data to identify the sentiments expressed is a dreary task. It is furthermore not an easy task to be performed by a computer system since the cues are very subtle to extract the sentiments.

A. Dataset And Methodology

The dataset was collected to analyse the sentiments on nuclear energy and global warming. The data is from Twitter that has 3000 tweets on nuclear energy. The data is scraped from Twitter. There are 38 attributes in the dataset and few are id, conversation_id, name, place, url, replies, photos, link, hashtag, retweet retweet_id etc. The sample row of data is shown below:

“1271534446345420802,1271534446345420802,15919924 10000,2020-06-12,16:06:50,EDT,868678380,holtecintl,Holt ec International,,"Holtec Highlights: Portable Robot Successfully Deployed to Make Multi-Purpose Canister Lid Closure Welds #nuclear #NuclearEnergy #manufacturers #Invention”

The dataset is imbalanced because it has more number of neutral sentiments (70%) than positive or negative. The neutral sentiment is assigned due to hesitant opinions. Analysis from such datasets becomes complicated. To resolve the complication HFS is used involving multiple experts opinions to find the alternative that is closest to the relevant sentiment. The steps followed for text preprocessing are:

- Text cleaning removing url patterns, mentions etc
- Removal of punctuations, stop-words, numeric characters and symbols (lemmatization)
- Identification of word count or frequency of words

After preprocessing the significant features that are common in both Naïve Bayes and Support Vector Machine (SVM) algorithms are chosen.

The algorithm for creation HFS:

- HFSs are constructed associated with triangular membership function.
- The linguistic variables used are strong positive, positive, moderate positive, strong negative, negative, moderate negative and neutral.
- The sentiment words are chosen manually and enlisted the help of experts to evaluate the words

The words chosen for assessment are like pollution, economy, reduction, energy, prosperity, threat, phasing out, low emission, possible alternative, safe, clean-up, lower cost, crisis etc. The rating table, Table 1. given below is used by the experts for assessing.

Table 1. Expert’s Rank Table

Linguistic Variables	Range
Strong Positive	0.9-1
Positive	0.8-0.6
Moderate Positive	0.7-0.5
Strong Negative	0.4-0.5
Negative	0.2-0.4
Moderate Negative	0-0.2
Neutral	0.0

The multiple opinions from multiple experts are aggregated using the hybrid weighted average method. For example the

tweet “ RT @mention From WSMV: TVA nuke construction, rates on board agenda: Nuclear power plant safety and fossil fuel emissions low-carbon emission... {link}” has a neutral opinion. After aggregation the predicted value neutral is likely to change to positive.

IV. RESULTS AND DISCUSSION

The sentiment analysis model used SVM, hybrid technique combining SVM–HFS and using geometric average for aggregation and same SVM-HFS along with hybrid weighted average operator.

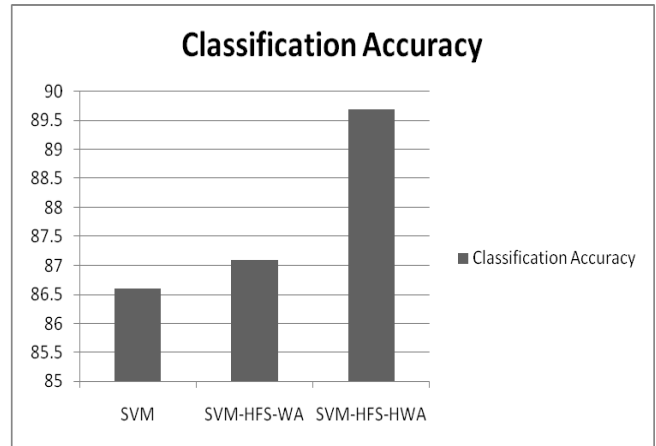


Figure 1. Sentiment analysis using HFS technique

Fig 1. shows the comparison graph on classification accuracy using SVM classifier combining HFS and different aggregation operators. Using hybrid weighted average method for aggregation has given better classification when compared to other techniques.

Table 2: Output showing negative/negative sentiment

```

Tweet1:hello japan is a nuclear power plant crisis. {link}- {'neg':
0.369, 'neu': 0.631, 'pos': 0.0, 'compound': -0.6249}
Tweet Positive Score :0.0
Tweet Negative Score :0.4
Tweet Neutral Score :0.1
Negative weight (wneg): 0.8
Neutral weight (wneu): 0.2
Positive weight (wpos): 0.0
Membership Functions Resultant :
St.Pos: [0. 0.2 0.2 0.2 0.2 0. 0. 0. 0. ]
M.Pos: [0. 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 ]
Pos: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. ]
St.Neg: [0. 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 ]
M.Neg: [0.8 0.8 0.6 0.4 0.2 0. 0. 0. 0. 0. ]
Neg: [0. 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 ]
Neu: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. ]
Aggregated Output: [0.8 0.8 0.6 0.4 0.2 0.2 0.2 0.2 0.2 0.2 ]
Defuzzified Output: 3.09
Output after Defuzzification: Negative
Sentiment : Negative
    
```

Table 2. gives a view of the tweet that is classified as negative. The sentiment/ opinions and the actual classified data are the same.

Table 3: Output showing positive/neutral sentiment

Tweet2 : renewable energy consumption tops nuclear for firsttime {link} is nuclear energy renewable {'neg': 0.0, 'neu': 0.545, 'pos': 0.455, 'compound': 0.7579}
 Tweet Positive Score :0.3
 Tweet Negative Score :0.0
 Tweet Neutral Score: 0.5
 Negative weight (w_{neg}): 0.0
 Neutral weight (w_{neu}): 0.0
 Positive weight (w_{pos}): 0.1
 Membership Functions Resultant :
 St.Pos: [0. 6. 0.6 0.2 0.6 0. 0. 0. 0. 0.]
 M.Pos: [0. 0.2 0.2 0.2 0. 0.2 0. 0. 0. 0.2]
 Pos: [0.6 0.6 0.4 0. 0. 0. 0. 0. 0. 0.]
 St.Neg: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 M.Neg: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 Neg: [0. 0.2 0. 0.2 0. 0. 0. 0. 0. 0.]
 Neu: [0. 0.4 0. 0.4 0. 0.2 0. 0. 0. 0.]
 Aggregated Output: [0. 0. 0. 0. 0. 0. 0.2 0.4 0.6 0.8]
 Defuzzified Output: 7.67
 Output after Defuzzification: Positive
 Sentiment: Neutral / author is just sharing information

Table 3. indicate that the tweet that is actually classified as neutral by the machine learning algorithm. But the expert opinions signify the other sentiment. The tweet in fact tells that nuclear energy is being considered as a renewable resource indicating a positive sentiment. There is subtle difference between the first statement in the tweet saying “energy consumption” and the second statement “renewable resource”. Thus the machine learning algorithm classifies the tweet as neutral. The weights computed based on experts’ suggestions has elevated the tweet from neutral to positive sentiment.

Table 4: Output showing negative/neutral sentiment

Tweet3: @mention If McGuinty has his way natural gas will be redundant and expensive wind, solar and dangerous nuclear energy will replace it.{link}- {'neg': 0.369, 'neu': 0.61, 'pos': 0.0, 'compound': -0.62}
 Tweet Positive Score :0.0
 Tweet Negative Score :0.4
 Tweet Neutral Score :0.1
 Negative weight(w_{neg}): 0.3
 Neutral weight (w_{neu}): 0.6
 Positive weight (w_{pos}): 0.1
 Membership Functions Resultant :
 St.Pos: [0. 0.2 0.2 0.2 0.2 0. 0. 0. 0. 0.]
 M.Pos: [0. 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2]
 Pos: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 St.Neg: [0. 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2]
 M.Neg: [0.6 0.6 0.6 0.4 0. 0. 0. 0. 0. 0.]
 Neg: [0. 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2]
 Neu: [0.4 0.4 0.2 0.2 0. 0.2 0. 0.2 0. 0.]
 Aggregated Output: [0.8 0.8 0.6 0.4 0.2 0.2 0.2 0.2 0.2 0.2]
 Defuzzified Output: 5.95
 Output after Defuzzification: Negative
 Sentiment : Neutral

Table 4. indicate that, similar to Table 3. the tweet that is actually classified is neutral by the machine learning algorithm. But the expert opinions signify the other sentiment. The tweet in fact tells that nuclear energy is being considered as a dangerous resource indicating a negative sentiment. Even in this tweet there is subtle difference between the first statement in the tweet saying “natural gas

being expensive” and the second statement “dangerous nuclear energy”. Thus the machine learning algorithm classifies the tweet as neutral. Using the weights computed values based on experts’ suggestions the tweet is classified from neutral to negative sentiment.

Table 5: Output showing neutral/neutral sentiment

Tweet4: mention - she is a mess. i am thinking of moving back down that way after 2012. here is another we did: {link} {'neg': 0.111, 'neu': 0.889, 'pos': 0.0, 'compound': -0.3612}
 Tweet Positive Score :0.1
 Tweet Negative Score :0.0
 Tweet Neutral Score: 0.6
 Negative weight (w_{neg}): 0.0
 Neutral weight (w_{neu}): 0.5
 Positive weight (w_{pos}): 0.1
 Membership Functions Resultant :
 St.Pos: [0. 0. 0.0 0.0 0.0 0. 0. 0. 0. 0.]
 M.Pos: [0. 0. 0. 0. 0. 0.0 0. 0. 0. 0.]
 Pos: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 St.Neg: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 M.Neg: [0.1 0. 3. 0. 0. 0. 0. 0. 0. 0.]
 Neg: [0. 0.0 0. 0.0 0. 0. 0. 0. 0. 0.]
 Neu: [0. 0.4 0. 0.6 0. 0.2 0. 0. 0.6 0.]
 Aggregated Output: [0.2 0.2 0.4 0.6 0.8 0.8 0.8 0.6 0.4 0.2]
 Defuzzified Output: 2.81
 Output after Defuzzification: Neutral
 Experts’ Sentiment: Neutral

Table 5. gives a view of the tweet that is classified as neutral. The actual opinion and the machine classification data are the same. The tweet is irrelevant o the subject of interest having no words related to nuclear energy. The frequency of words related to nuclear energy is less, hence showing neutral review.

The values after defuzzification determine the category or the class the tweet belongs. The value between 7 and 10 is treated as positive, value between 5 and 7 as negative and below range 5 is considered to be neutral.

V. CONCLUSION

As the number of product reviews in online grows, more attention is being paid to the method of using these reviews to assist in decision making. This study presents new ways to take full advantage of HFS theory, Sentiment Analysis, and Machine learning to solve product hesitant opinions. Python is used to crawl the related reviews of various features from kaggle websites. The sentiment score for each feature based on the proposed framework is computed. This gives an overall performance of for each feature of each tweet. .Based on a developed fuzzy sentiment word framework and use of HFS and aggregation operators, the hybrid method of averaging the features computes a better sentiment score for each feature word. There is an improved distinction among the HFS method and baseline approaches that helps to balance the dataset with more neutral sentiments labeled.

REFERENCES

1. Atanassov, Krassimir. "Intuitionistic fuzzy sets." International Journal Bioautomation 20 (2016): 1.



2. Liao, Huchang, et al. "Hesitant fuzzy linguistic term set and its application in decision making: a state-of-the-art survey." International Journal of Fuzzy Systems 20.7 (2018): 2084-2110.
3. Torra, V.: Hesitant fuzzy sets. Int. J. Intell. Syst. 25(6), 529-539,(2010)
4. Rodríguez, Rosa M., et al. "Hesitant fuzzy sets: state of the art and future directions." International Journal of Intelligent Systems 29.6 (2014): 495-524.
5. 5.N. Chen, Z.S. Xu, and M.M. Xia. Correlation coefficients of hesitant fuzzy sets and their applications to clustering analysis. Applied Mathematical Modelling, 37(4):2197-2211, 2013.
6. Sahar Sardari, Mahdi Eftekhari, Fatemeh Afsari, Hesitant fuzzy decision tree approach for highly imbalanced data classification, Applied Soft Computing, Volume 61, 2017, Pages 727-741, ISSN 1568-4946
7. Ebrahimpour, Mohammad Kazem, and Mahdi Eftekhari. "Ensemble of feature selection methods: A hesitant fuzzy sets approach." Applied Soft Computing 50 (2017): 300-312.
8. Qin, Y., Wang, X. & Xu, Z. Ranking Tourist Attractions through Online Reviews: A Novel Method with Intuitionistic and Hesitant Fuzzy Information Based on Sentiment Analysis. Int. J. Fuzzy Syst. (2021)
9. Dutta, Palash and Rupjit Saikia. "A Decision-making Approach for Choosing a Reliable Product under the Hesitant Fuzzy Environment via a Novel Distance Measure." Vikalpa: The Journal for Decision Makers 45 (2020): 147 - 159.
10. Xia, Meimei, Zeshui Xu, and Na Chen. "Some hesitant fuzzy aggregation operators with their application in group decision making." Group Decision and Negotiation 22.2 (2013): 259-279.
11. Meimei Xia, Zeshui Xu, Hesitant fuzzy information aggregation in decision making, International Journal of Approximate Reasoning, Volume 52, Issue 3, 2011, 395-407

AUTHORS PROFILE



Dr. Roopa Chandrika R Professor Department of CSE-AI&ML Malla Reddy College of Engineering and Technology, MRCET Campus Hyderabad, India, Email: mroopachandrika@gmail.com.



Dr. N.S. Gowri Ganesh Professor Department of Information Technology Malla Reddy College of Engineering and Technology, MRCET Campus Hyderabad, India, Email: gowriganesh@gmail.com .



Dr. A. Mummoorthy, Professor Department of Information Technology Malla Reddy College of Engineering and Technology, MRCET Campus Hyderabad, India, Email: amummoorthy@gmail.com.



Gayathri M, Associate Pofessor Department of CSE-AI&ML Malla Reddy College of Engineering and Technology, MRCET Campus Hyderabad, India, Email: gayatricse312@gmail.com.