2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT) | 978-1-6654-1924-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/WIIAT50758.2020.00149

Integration of Fuzzy and LSTM in Three-Way Decisions

L.D.C.S. Subhashini[1,2,3,4], Yuefeng Li* [1,5], Jinglan Zhang[1,6], Ajantha S. Atukorale[2,7],
[1] School of Computer Science, Queensland University of Technology, Australia
[2]University of Colombo, Sri Lanka
[3]University of Sri Jayewardenepura, Sri Lanka
[4]subhashini@sjp.ac.lk, shashikala.charles@hdr.qut.edu.au
[5]y2.li@qut.edu.au
[6]jinglan.zhang@qut.edu.au
[7]aja@ucsc.cmb.ac.lk

***Abstract*—Recently, a three-way decision concept is studied to handle the uncertainty and incompleteness in the given attribute set based on acceptance, rejection, and uncertain regions. A three-way decision is a vital method in solving the problem of uncertainty in opinion mining models.There are models developed to solve the problem of uncertainty under three-way decision theory. However, with the classification performance of existing models, we further investigate the formulation of three-way regions using fuzzy concepts and Long Short Term Memory (LSTM). To demonstrate the effectiveness of the method, experiments were conducted using a movie review and an ebook review datasets. The experimental results show that the proposed three-way framework is effective for dealing with uncertainties in opinions.**

***Index Terms*—Fuzzy Concept, Three-way Decision, LSTM**

I. Introduction

The three-way decision, as expressed by the positive, boundary, and negative rules, reflects more accurately the philosophy. It focuses on the actions implied by decision rules, rather than their statistical features [1]. As a novel methodology to deal with uncertain and incomplete data, the three-way decision has attracted much attention in the area of decision-making in opinion mining [2]. The concept of uncertainty refers to unclear and inconsistent opinions. Therefore, the ability to handle uncertainty is essential in opinion classification models [3]. When it comes to opinion classification models, the mechanism for managing these uncertainties is a challenging task as many user generated opinions contain uncertainties [4]. A fundamental issue in applying three-way decision is how to construct three regions with reasonable interpretation. There are several three-way models developed by researchers to solve the problem of uncertainity [5]–[8].

Studies have shown that classifying the uncertainty boundary is essential. Fuzzy logic has been used by researchers to solve the problem of uncertainties in opinion mining [2], [3], [9]–[11]. It represents uncertainties of features using fuzzy-values between zero and one. There is a is a critical problem for using fuzzy concepts to make binary decisions when the fuzzy similarity values to categories are very close [2], [11]. They found that there are a few errors in POS and NEG, but there are many errors in BND. Therefore, the tough research issue is how to find BND and then further classify it into POS and NEG regions correctly [2]. We call the three regions the positive region (POS), negative region (NEG) and uncertain boundary (BND), respectively.

The above studies show that the classification of the boundary region is essential in three-way decision. We attempt to solve this challenging issue via Long Short Term Memory (LSTM). The proposed method classifies reviews into three regions in the first stage using fuzzy composition: the positive region, negative region and uncertain boundary. It then splits the uncertain boundary into positive and negative regions using LSTM. The three-way decision framework provides an elegant way to integrate fuzzy logic and LSTM together. The popular way to find more specific features is to use three feature selection indexes [2]. It firstly discovered closed patterns using the selected features, and then selected fuzzy formal concepts (higher-level concepts) from the discovered closed patterns. The selected concepts are also used for the three-way decision framework to classify reviews into three regions in the first stage. After that, in the second stage, we integrate word-embedding vectors with the selected concepts and identify the difference between vectors and their mean value. The updated feature vectors are then used for a LSTM algorithm to split the boundary region clearly into negative and positive regions. The proposed framework provides valuable insights for dealing with uncertainties in user opinions. The main contribution is an effective way to classify uncertain boundaries in three-way decisions using LSTM.

II. Related Work

Three-way decision theory allows a risk-based way to understand the lower and upper approximation of the probabilistic positive, boundary and negative region. The three-way decisions are formally described in an approximation space where the probabilistic theory is used to measure lower and upper approximations for risk-based decision making [4]. The

*Corresponding Author

978-1-6654-1924-6/20/$31.00 ©2020 IEEE
DOI 10.1109/WIIAT50758.2020.00149

three regions enable us to derive three types of decision rules: positive rules for acceptance, boundary rules for indecision or delayed decision, and negative rules for rejection [1]. Three-way concept lattice is studied to handle the uncertainty and incompleteness in the given attribute set based on acceptance, rejection, and uncertain regions [5], [6]. There is a challenge, how to reflect the risk attitude in deter-mining decision rules. A new model is introduced to use prospect theory into three-way decision to construct a novel three-way decision model [12]. A three-way two-stage decision model is proposed based on semantic features for Chinese irony detection [7]. In the first stage, a classifier is trained to predict the ironic tendency based on the lexical features. Then in the second stage, new features are merged with five features. It is a stiff challenge to apply in image data. The problem is feature extraction method [13]. A deep neural network is used to identify features. As an uncertain decision approach the three-way decision has developed rapidly in the fields of decision-making, granular computing, and incomplete data analysis [8]. A model is proposed considering the consideration of uncertain and multi-level characteristics of recommendation information. It combines three-way decisions with granular computing, and builds a novel dynamic three-way recommendation model to address the limitations of static two-way recommendation [14].

Pattern-based methods are also used as a higher level method to discover knowledge from text data [15], [16]. Pattern mining can discover sequencing terms that frequently co-occur in a customer review, and as such, a set of terms can effectively represent the knowledge of reviews. Frequent patterns and closed patterns are frequently employed to represent knowledge and trends in a dataset [15], [17].

The fuzzy composition is used to calculate relevance fuzzy values for each category [2], [3], [9], [10]. These researchers have used fuzzy composition to decide the relevance category. However, when the fuzzy values are close to each other, it is hard to decide the relevance category. In Li et al. model [3], when the fuzzy relevance values were close to each other it is difficult to classify opinions into relevant categories. Therefore, we attempted to solve the problem by integrating fuzzy concepts and deep learning in a framework [2].

LSTM employs a bidirectional variant to capture sentiment for sentiment classification [18], [19]. The model has shown a 77.4% percentage of F-measure in the Movie review dataset [20]. Recursive Tensor Neural Network is used to model correlations between different child node vector dimensions, showing 75.9% F-measure on the movie review dataset [21] .In LSTM first, words in posts are converted into vectors using word embedding models. Then, the word sequence in sentences are input to LSTM to learn the long distance contextual dependency among words [22].

III. Proposed Model

In the three-way decision framework it essential to find a way to classify uncertain boundaries. We firstly used the fuzzy-based model to infer the positive, negative and uncertain boundary. After that, LSTM was used to classify uncertain boundaries.

A. Fuzzy-based concept discovery

Term-based feature selection approaches are suffering from polysemy and synonymy [23] when using them for dealing with uncertainties within user reviews. A popular way for solving this problem in text mining is to extend low-level term spaces to higher level patterns or concepts [15]. In the following sections, we firstly generate formal concepts within user reviews. We then provide a method to weight associations between concepts and selected term features. At last, we discuss how to use fuzzy composition to make a classification in three-way.

1) Feature selection for generating formal concepts: Formal concepts have elegant properties; however, closed patterns only discuss terms' binary appearances and long closed patterns have very low frequency [24]. In this paper, we select some relevant term features to solve these issues first to reduce the time complexity for finding concepts, then using fuzzy composition to find the associations between concepts and categories.

In this research, reviews are pre-processed and applied BM25, Uni, and ICF (three feature selection techniques) to find the useful terms. This is a significant step as we need to conduct many experiments to decide the combination strategy of using multiple feature selections. At present, as identified in the literature review, there are several techniques available for pre-processing. This process is very much vital for improving the performance of the classification process of customer reviews. Natural Language Processing (NLP) techniques of stemming, stop words removal and tokenization for pre-processing were applied. They have been identified by literature as the best techniques are presently available. The output will be a set of terms, as shown in Table I.

TABLE I
REVIEWS AFTER FEATURE SELECTION

Review	funny	good	pretty	great	comedy	awful
r1	x		x			
r2	x	x		x		
r3	x		x			
r4		x		x		x
r5	x		x		x	
r6	x	x		x		
r7	x	x				x

Feature selection indexes of BM25, Uni, and ICF are applied for the term selected. We have applied feature selection separately for both categories with the same method. In this model multiple indexes of $BM25>30$, Uni $>$ 0.2 and $ICF<\log(2)$ are used. The condition of $BM25>30$ was suggested based on the experiment results and the other two were based on the past literature [3]. The three conditions were checked simultaneously and if one condition did not satisfy, the term is eliminated. Table I shows an example of feature selection, where we assume that term set $T =$

$\{funny, good, pretty, great, comedy, awful\}$ and review set $R = \{r_1, r_2, ..., r_7\}$.

The frequency of terms might vary according to user reviews' size. The normalization values of BM25 weights are calculated in Eq. 1 [25] for a given category j (for binary classification, j is either the positive class or negative class).

$$nBM25(t_i) = BM25(t_i)\log_2(1 + c\frac{avgR}{df_{ij}}) \qquad (1)$$

where $avgR$ is the average review length in the whole collection; t_i is the original term frequency; df_{ij} is the number of reviews in term i in category j and c is the free parameter of the normalization method which is 0.75. We have decided the value based on our experiments.

The next task is to define a threshold to reduce the number of noisy terms. The significance of the feature selection method is to identify relevant features for mining closed patterns, and also reduce the size of reviews for generating closed patterns efficiently.

Closed patterns were generated using frequent itemset generation in FTree algorithm by defining a minimum support [26]. A frequent pattern is considered as closed pattern if $X = Cls(X)$ [27].For a given set of reviews Y, we can define their termset (a set of terms that are used to describe these reviews) as follows (in Eq.2):

$$termset(Y) = \{t | \forall r \in Y \Longrightarrow t \in r\}. \qquad (2)$$

Then we can define the closure of termset X (in Eq.3) using the above definitions.

$$Cls(X) = termset(coverset(X)). \qquad (3)$$

Identified patterns are larger than the terms. Hence, how to effectively deal with large amount of discovered patterns was the next challenge. Closed pattern mining algorithm with a minimum support value which is greater than 20 was introduced in our experiments. Thereafter, the change of the minimum support is used to control the size of discovered patterns. The discovered closed patterns are further processed to determine the formal concepts using a suitable minimum support as the threshold. Table II shows the selected formal concepts from closed patterns $CP = \{p_1, p_2, p_3, p_4, p_5\}$ based on the threshold value.

TABLE II
SELECTED FORMAL CONCEPTS

Concept	Intent	Extent	Original Pattern
c1	funny, pretty	r1, r3, r5	p1
c2	funny, good	r2, r6, r7	p5

B. Three-way Classification

We can find a set of concepts C for a given set of reviews R, where R is either a set of positive reviews or a set of negative reviews, and we have

$$C = \{c = (Terms(p), coverset(p)) | p \in CP, supp_r(p) \geq \theta\}$$

where parameter θ is a minimum support; $Terms(p)$, the set of terms used in pattern p, is the intent of c and $coverset(p)$ is the extent of c.

In this sub-section, we discuss the relation between concepts and categories (e.g., the positive class and negative class for binary classification) using fuzzy composition [3] in order to classify reviews into three regions: the positive, negative and uncertain boundary.

The relation between concepts and categories is denoted as (I_{C-Catg}) which can be evaluated by using the fuzzy composition to integrate relation (I_{C-T}) between concepts and terms and relation (I_{T-Catg}) between terms and categories as described in Eq. 4.

$$I_{C-Catg} = I_{C-T} \circ I_{T-Catg} \qquad (4)$$

where $I_{C-Catg}(i,j) = max_{k \in T} min[I_{C-T}(i,k), I_{T-Catg}(k,j)]$.

The I_{C-T} relation is used to describe the uncertain factor of terms in a concept. In this paper, we use the normalized BM25 value $nBM25(t_k)$ as the uncertain factor for term k in concept i (see Eq. 5).

$$I_{C-T}[i,k] = nBM25(t_k) \qquad (5)$$

The relation of Term-Category (I_{T-Catg}) can be obtained using the fuzzy composition over the Term-Review relation (I_{T-R}) and Review-Category relation (I_{R-Catg}) as shown in Eq. 6. In this paper, we use $nBM25(t_k)$ values to represent the relevant strength between terms and reviews and reviews and categories, respectively. Therefore (see the Eq. 7), $I_{T-Catg}(k,j)$ can be calculated using $nBM25$ weights for a given category j.

$$I_{T-Catg} = I_{T-R} \circ I_{R-Catg} \qquad (6)$$

$$I_{T-Catg}(k,j) = \begin{cases} nBM25(t_k), & \text{if } nBM25(t_k) < 1 \\ 1 & \text{otherwise} \end{cases} \qquad (7)$$

Algorithm 1 illustrates the idea for classifying user reviews into three regions: the positive region (POS), negative region (NEG) and the uncertain boundary (BND). It firstly describes the training process (the first for loop) for calculate the relation I_{C-Catg} (a concept-category matrix) by using the fuzzy composition. It then calculates a fuzzy value for each new review r to a category j (see the second for loop). At last, it determines the three regions (POS, NEG and BND) based on these fuzzy values for a given unlabeled review set U.

Algorithm 2 uses a parameter δ (a minimal value) to classify reviews into three regions. It is challenging to decide the class of a review when its fuzzy values to the two classes are very closed [2].

We have used a probability based loss function to decide δ [2], [4], in this paper we use a standard derivation to decide parameter δ as it is a very small value. Algorithm 2 describes the process for calculating the value for parameter δ based on a training set R and the concept-term relations. The time complexity of Algorithm 2 is $O(|T| * |C| * |R|)$. As the numbers of terms and concepts are not very big, the algorithm for deciding δ is efficient.

Algorithm 1 Three-way Classification

Require: C, I_{C-T}, I_{T-Catg}, U, two categories: $j = 1$ and $j = 0$ and an experimental coefficient δ
Ensure: I_{C-Catg} and three regions: POS, NEG, BND
for each j **do**
  for $c_i \in C$ do **do**
    $I_{C-Catg}(i, j) =$
    $max_{k \in T} min[I_{C-T}(i, k), I_{T-Catg}(k, j)]$
  end for
end for
for $r \in U$ **do**
  $C_r = \{c \in C | intent(c) \subseteq r\}$
  for j do **do**
    $f_j(r) = max_{c_i \in C_r}[I_{C-Catg}(i, j)]$
  end for
end for
for $r \in U$ **do**
  $POS = \{r \in R, f_{j=1}(r) - f_{j=0}(r) > \delta\}$
  $NEG = \{r \in R, f_{j=0}(r) - f_{j=1}(r) > \delta\}$
  $BND = \{r \in R, |f_{j=0}(r) - f_{j=1}(r)| \leq \delta\}$
end for

Algorithm 2 Deciding parameter δ

Require: R, I_{C-T} two categories: $j = 1$ and $j = 0$
Ensure: δ
for $r \in R$ **do**
  for each j **do**
    $f_j(r) = \sum_{t_k \in r, intent(c_i) \subseteq r} I_{C-T}(i, k)$
  end for
  $f(r) = \frac{f_0(r) + f_1(r)}{2}$
end for
for $r \in R$ do **do**
  let $\mu = \frac{1}{|R|} \sum_{r \in R} f(r)$
end for
let $\delta = \sqrt{\frac{\sum_{r \in R}(f(r) - \mu)^2}{|R| - 1}}$

It shows that the uncertain boundary (BND) includes a lot of uncertain reviews. Table III and Table IV show the size and the percentage of the uncertain boundary in both movie review and ebook review datasets, respectively.

TABLE III
BOUNDARY REVIEWS-MOVIE REVIEW DATASET

Category	Boundary Reviews	Precentage
Positive	115	11.5%
Negative	102	10.2%
Total	217	21.7%

TABLE IV
BOUNDARY REVIEWS-EBOOK REVIEW DATASET

Category	Boundary Reviews	Precentage
Positive	104	10.4%
Negative	101	10.1%
Total	205	20.5%

1) Boundary Classification using LSTM: We further classify the uncertain boundary (BND) as the data in BND are not linearly separable. In this paper, we assume that it may be possible to map the data into a higher dimensional space resulting in a linearly separable set. There are many methods to do the mapping. For example, we may map a set of terms (a n-dimensional vector) $[t_1, t_2, ..., t_n]$ to a 2n-dimensional vector $[t_1, t_2, ..., t_n, t_1^2, t_2^2, ..., t_n^2]$. However, it is likely to find the dependency between features. In this paper we use word embedding to do the mapping. It can find a higher dimensional space easily. Word embedding models map each word from the vocabulary to a vector of real numbers. We use word2vec model [28] in our experiments and each word was encoded by 12-dimention vector, that is,

$$\overrightarrow{w} = (x_1, x_2, ..., x_12)$$

for all words $w \in \Omega$, where x_i are real numbers. The number of dimension was decided based on our experiments.

To solve the hard issue, in this paper, we integrate word embedding vectors and the terms (with assigned fuzzy values) that we selected from the intents of the formal concepts; and then use a LSTM classifier to classify the uncertain boundary.

Algorithm 3 describes the new idea for integrating word embedding vectors and the terms (with assigned fuzzy values) in the formal concepts. The first for loop firstly gets the related concepts of each word $w \in \Omega$ for the positive class j. It then uses the average fuzzy value to update the word vector, where $\overrightarrow{e_i}$ are unit vectors. In the second for loop it uses Z-core normalization to normalize the word vectors by using the average vector (μ) and the standard deviation s. The time complexity of Algorithm 3 is $O(|\Omega| * |C|)$ that is decided by the first for loop. In the LSTM architecture, there are three gates and a cell memory state.After representing each word by its corresponding vector trained by above word embedding model, the sequence of words $t_1, , t_n$ are input to LSTM one by one in a sequence as in Figure 1. In the first stage words $t_1, , t_n$ are converted to $x_1, , x_n$ using above trained word embedding model. At each time j, the output W of the hidden layer H_n will be propagated back to the hidden layer together with the next input x_n+1 at the next point of time $j + 1$. Finally, the last output W_n will be fed to the output layer.

To generate LSTM gates, we convert the post into three dimensions as dimensions of word embedding, the number of words in the review and the number of reviews. In this model we adapt to a single hidden layer neural network. The number of neurons in input layer is the same as the dimension of word embedding model, and the number of neurons in the output layer one to define two classes.

We have applied gradient-based back propagation and adjust the weight in hidden layer at each point in time. After several 100 epochs of training, we can obtain the sentiment classification model for boundary region with the following F-measure as in Figure 2 .

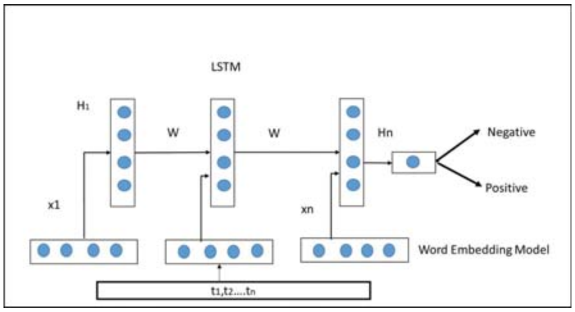

Fig. 1. LSTM for Boundry Classification

Algorithm 3 New Word Embedding Vector for LSTM

Require: Ω is a set of words, each word $w \in \Omega$ is a word2vector $\overrightarrow{w}$; and C is the concept set

Ensure: new word vectors $\hat{\overrightarrow{w}}$ for all words $w \in \Omega$.

for $w \in \Omega$ do **do**
  let $C_w = \{c \in C | w \in intent(c)\}$ and $j = 1$
  $\overrightarrow{w} = \overrightarrow{w} + \sum_{i=1}^{8} (\frac{\sum_{c_i \in C_w} I_{C-Catg}(i,j)}{|C_w|} \overrightarrow{e_i})$
end for
for $w \in \Omega$ do **do**
  let $\overrightarrow{\mu} = \frac{\sum_{w \in \Omega} \overrightarrow{w}}{|\Omega|}$
  let $s = \sqrt{\frac{1}{|\Omega|} ||\overrightarrow{w} - \overrightarrow{\mu}||}$
  $\hat{\overrightarrow{w}} = \frac{1}{s}(\overrightarrow{w} - \overrightarrow{\mu})$
end for

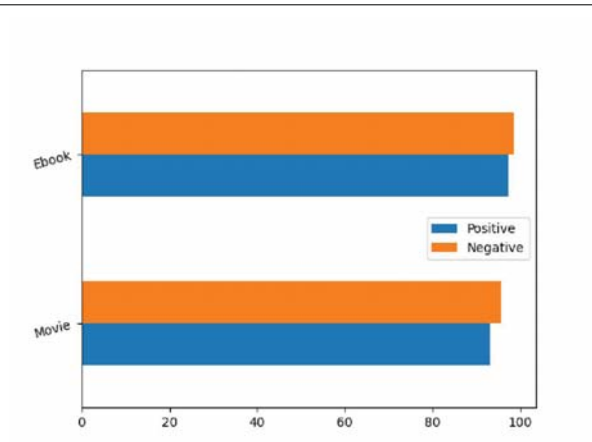

Fig. 2. F-measure of Boundary Classification

IV. Experiment Results

Many experiments were conducted to evaluate the effectiveness of the proposed model. The purpose of these experiments is to evaluate our model and compare it with other state-of-the-art classification models. The performance of the model is measured by the F-measure. Also, a pair wise t-test is performed to verify the significance.

A. DataSets

The movie review dataset and ebook review dataset is used for the evaluation. These datasets are very popular in opinion mining [3], [29].The movie review dataset and ebook review dataset consists of 2000 movie reviews, with 1000 reviews for each category (negative and positive) for both training and testing datasets.

B. Baseline Models

This research provides a comprehensive evaluation of the proposed model, we have selected five baseline models.

- Li and Tsai's [3] fuzzy model was developed and evaluated on movie review and ebook review datasets.
- Support Vector Machine(SVM) is used since it is outperformed model in opinion mining in the literature [3].
- CNN model of Kim [30] is a deep learning model for opinion mining.
- Graph Convolutional Networks (GCN) [31] is a recent model developed using GCN and has shown significant. results.
- Attention Network [32] used Multi-sentiment-resource Enhanced Attention Network (MEAN) to alleviate the problem by integrating three kinds of sentiment linguistic knowledge.

C. Results

In order to evaluate the effectiveness of the proposed model, we compare the results with existing classification models as in Table V and Table VI.

TABLE V
F-Measure of State-of-the-Art Models: Movie Review

Model	Precision	Recall	F-measure
Proposed Model	0.9234	0.9352	**0.9292**
FFCM	0.8870	0.8840	0.8800
CNN	Not Given	Not Given	0.8150
SVM	0.8770	0.8690	0.8730
GCN	Not mention	Not mention	0.7674
Attention	Not mention	Not mention	0.8450

It is obvious that the proposed model has the best performance in the two datasets comparing with models: FFCM [3], CNN [30], SVM [3], SVM [29], GCN [31] and Attention Network [32]. In Table VI we have not included CNN [30], GCN [31] and Attention Network [32] because those models were not evaluated with ebook review dataset.

We also conducted the statistical significance testing (a two-tailed t-test). The t-test results show that the proposed model is better than the five baseline models. It has shown from Table VII our model is more significant than CNN and Attention Network which indicated that increment is > 0.5.

TABLE VI
F-MEASURE OF STATE-OF-THE-ART MODELS: EBOOK REVIEW

Model	Precision	Recall	F-measure
Proposed Model	0.9657	0.9678	**0.9667**
FFCM	0.9509	0.9508	0.9509
SVM	0.9382	0.9381	0.9382

TABLE VII
P-VALUES-TWO TAILED

Model	Movie	Ebook
FFCM	0.0526	0.0423
CNN	0.0432	-
SVM	0.0356	0.0265
GCN	0.0187	-
Attention	0.0234	-

V. CONCLUSION

This paper focused on three-way decisions using fuzzy concepts and LSTM. A method is proposed to generate three regions positive, negative and boundary region using fuzzy concepts. Thereafter, the boundary region is classified using LSTM. This framework addresses the problem of uncertainties in opinion classification, and provides a promising way to construct an opinion classifier for boundary region. The experiment results show that the model can significantly improve the performance of binary opinion classification.

ACKNOWLEDGMENT

This research was supported by the Queensland University of Technology (QUT), Australia, University Grant Commission (UGC), Sri Lanka and University of Colombo School of Computing (UCSC), Sri Lanka.

REFERENCES

[1] Y. Yao, "Three-way decision: an interpretation of rules in rough set theory," in *International Conference on Rough Sets and Knowledge Technology*. Springer, 2009, pp. 642–649.

[2] L. D. C. S. Subhashini, Y. Li, J. Zhang, and A. S. Atukorale, "Integration of fuzzy and deep learning in three-way decisions." *International Conference on Data Mining Workshop*, 2020.

[3] S.-T. Li and F.-C. Tsai, "A fuzzy conceptualization model for text mining with application in opinion polarity classification," *Knowledge-Based Systems*, vol. 39, pp. 23–33, 2013.

[4] Y. Li, L. Zhang, Y. Xu, Y. Yao, R. Y. K. Lau, and Y. Wu, "Enhancing binary classification by modeling uncertain boundary in three-way decisions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 7, pp. 1438–1451, 2017.

[5] P. K. Singh, "Three-way fuzzy concept lattice representation using neutrosophic set," *International Journal of Machine Learning and Cybernetics*, vol. 8, no. 1, pp. 69–79, 2017.

[6] Y. Yao, "Three-way granular computing, rough sets, and formal concept analysis," *International Journal of Approximate Reasoning*, vol. 116, pp. 106–125, 2020.

[7] X. Jia, Z. Deng, F. Min, and D. Liu, "Three-way decisions based feature fusion for chinese irony detection," *International Journal of Approximate Reasoning*, vol. 113, pp. 324–335, 2019.

[8] T. Wang, H. Li, L. Zhang, X. Zhou, and B. Huang, "A three-way decision model based on cumulative prospect theory," *Information Sciences*, vol. 519, pp. 74–92, 2020.

[9] T. Quan, S. Hui, and T. Cao, "A fuzzy fca-based approach for citation-based document retrieval," in *2004 IEEE Conference on Cybernetics and Intelligent System*, vol. 1. IEEE, 2004, pp. 578–583.

[10] K. Ravi, V. Ravi, and P. S. R. K. Prasad, "Fuzzy formal concept analysis based opinion mining for crm in financial services," *Applied Soft Computing*, vol. 60, pp. 786–807, 2017.

[11] L. D. C. S. Subhashini, Y. Li, J. Zhang, and A. S. Atukorale, "Three-way framework using fuzzy concepts and semantic rules in opinion classification," *Australasian Joint Conference on Artificial Intelligence*, 2020.

[12] T. Wang, H. Li, X. Zhou, B. Huang, and H. Zhu, "A prospect theory-based three-way decision model," *Knowledge-Based Systems*, p. 106129, 2020.

[13] H. Li, L. Zhang, X. Zhou, and B. Huang, "Cost-sensitive sequential three-way decision modeling using a deep neural network," *International Journal of Approximate Reasoning*, vol. 85, pp. 68–78, 2017.

[14] D. Liu and X. Ye, "A matrix factorization based dynamic granularity recommendation with three-way decisions," *Knowledge-Based Systems*, vol. 191, p. 105243, 2020.

[15] Y. Li, A. Algarni, M. Albathan, Y. Shen, and M. A. Bijaksana, "Relevance feature discovery for text mining," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 6, pp. 1656–1669, 2015.

[16] Y. Li, A. Algarni, and N. Zhong, "Mining positive and negative patterns for relevance feature discovery," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 753–762.

[17] N. Zhong, Y. Li, and S.-T. Wu, "Effective pattern discovery for text mining," *IEEE transactions on knowledge and data engineering*, vol. 24, no. 1, pp. 30–44, 2012.

[18] M. Huang, Y. Cao, and C. Dong, "Modeling rich contexts for sentiment classification with lstm," *arXiv preprint arXiv:1605.01478*, 2016.

[19] S. Poria, E. Cambria, G. Winterstein, and G.-B. Huang, "Sentic patterns: Dependency-based rules for concept-level sentiment analysis," *Knowledge-Based Systems*, vol. 69, pp. 45–63, 2014.

[20] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.

[21] R. Socher, A. Perelygin, J. Wu, J. Chuang, C. D. Manning, A. Ng, and C. Potts, "Recursive deep models for semantic compositionality over a sentiment treebank," in *Proceedings of the 2013 conference on empirical methods in natural language processing*, 2013, pp. 1631–1642.

[22] J.-H. Wang, T.-W. Liu, X. Luo, and L. Wang, "An lstm approach to short text sentiment classification with word embeddings," in *Proceedings of the 30th conference on computational linguistics and speech processing (ROCLING 2018)*, 2018, pp. 214–223.

[23] Y. Li, A. Algarni, and N. Zhong, "Mining positive and negative patterns for relevance feature discovery," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 753–762.

[24] N. Zhong, Y. Li, and S.-T. Wu, "Effective pattern discovery for text mining," *IEEE transactions on knowledge and data engineering*, vol. 24, no. 1, pp. 30–44, 2012.

[25] B. He and I. Ounis, "Term frequency normalisation tuning for bm25 and dfr models," in *European Conference on Information Retrieval*. Springer, 2005, pp. 200–214.

[26] M. Mahdi, S. Abdelrahman, R. Bahgat, and I. Ismail, "F-tree: an algorithm for clustering transactional data using frequency tree," *arXiv preprint arXiv:1705.00761*, 2017.

[27] A. Algarni, "Relevance feature discovery for text analysis," Ph.D. dissertation, Queensland University of Technology, 2011.

[28] Y. Goldberg and O. Levy, "word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method," *arXiv preprint arXiv:1402.3722*, 2014.

[29] R. Moraes, J. F. Valiati, and W. P. G. Neto, "Document-level sentiment classification: An empirical comparison between svm and ann," *Expert Systems with Applications*, vol. 40, no. 2, pp. 621–633, 2013.

[30] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.

[31] L. Yao, C. Mao, and Y. Luo, "Graph convolutional networks for text classification," *arXiv preprint arXiv:1809.05679*, 2018.

[32] Z. Lei, Y. Yang, M. Yang, and Y. Liu, "A multi-sentiment-resource enhanced attention network for sentiment classification," *arXiv preprint arXiv:1807.04990*, 2018.