A Deep Learning-Fuzzy Based Hybrid Ensemble Approach for Aspect Level Sentiment Classification

Tanu Sharma[1], Kamaldeep Kaur[2]
[1,2]University School of Information, Communication, & Technology, GGSIPU, Delhi, India
E-mail: tanu.phd.132.usict2018@ipu.ac.in, kdkaur99@ipu.ac.in

Keywords: aspect based sentiment analysis, fuzzy logic, ensemble, deep neural network, natural language processing

Received: January 10, 2023

Aspect level sentiment classification (ALSC) has gained high importance in the era of an e-commerce-based economy. It allows manufacturers to improve the designs of their products based on users' feedback. However, only a few datasets of limited domains are available for ALSC task. To push forward the research in automated ALSC, this study contributes car dataset of the automobile domain. In this study, a novel fuzzy ensemble technique is also proposed based on the mathematical analysis of confidence scores of base deep neural networks. The proposed approach allows for the correction of the misclassifications of base deep learners through a reward and penalization strategy. The experimental results on five benchmark datasets show that the proposed approach outperforms the constituent base deep neural networks and several other important baselines. The proposed Fuzzy ensemble also performed at par with the most recent Graph Convolution Neural Networks on the basis of Friedman and Nemenyi Tests.

Povzetek: V prispevku sta opisani dve novosti: (1) baza podatkov o avtomobilih na temo aspektnega nivoja sentimentne klasifikacije in (2) nova mehka ansambelska metoda za kombiniranje klasifikacij globokih nevronskih mrež.

1 Introduction

Aspect based sentiment analysis predicts the polarity of sentiment towards a specific entity or target, thus providing more detailed information as compared to general sentiment analysis. Aspect level sentiment classification (ALSC) specifically handles the sentiment classification task of ABSA [1]. Recently, the research in the field of ALSC is driven by well-performing deep learning methods. Most of the researchers are leveraging deep learning (DL) methods for achieving better accuracy on benchmark datasets. However, the benchmarking study [2] of the latest deep learning methods revealed that the good performance of deep learning methods is cursed with poor performance in terms of training time. Moreover, despite the effectiveness of the latest methods in ALSC, it is challenging to apply such methods to real-world applications because of the unavailability of labeled data in various domains.

So far, multiple datasets have been proposed in ALSC, that includes SemEval's restaurant14 [3],laptop14 [3], restaurant15 [4], restaurant16 [5], MAMS [6], and Twitter [7] dataset. Although these datasets are studied as benchmark datasets in almost every research of ALSC, these datasets lack domain diversity. Most of them belong to the restaurant domain except laptop14 and Twitter datasets. It is a well-known fact that supervised approaches like deep learning rely on properly labeled data for training. However, there is a lack of availability of datasets of different categories of products that conform to SemEval guidelines. Thus, the applicability of well-performing methods in product domains other than mobile phones and laptops is not well-tested and hence doubtful. To fill the gaps in the research domain of ALSC, this study provides a dataset in the domain of automobiles that conforms to SemEval guidelines. The availability of labeled data in the automobile domain will help the researchers and other stakeholders of the automobile domain to automate the ALSC process efficiently. Another advantage of proposing a dataset in the automobile domain is to facilitate cross-domain transfer learning in the field of ALSC.

Ensemble learning is a paradigm where decision scores from multiple base learners are collectively used to predict the outcome of a given input sample. An ensemble model aims to capture the salient features of the base methods and thus ensures providing promising results as compared to its base learners. The ensemble is constructed by taking prediction decisions from various base learners. For this purpose, some pre-defined weight is allocated to each contributing leaner and the outcome is calculated. However, these methods do not pay any attention to the confidence level of the prediction made by the base learners. Most of the ensemble techniques ignore the confidence score of the predictions made by the learner. In this study, the confidence score (probability) of the predictions of each base learner is considered and then this score value is utilized to calculate the final prediction for each sample.

The contributions of this study are twofold-

- A dataset in the automobile domain specifically for ALSC task is developed.
- An efficient ensemble technique for the ALSC task is proposed. To build an ensemble, a thorough investigation of the latest deep learning methods is performed to select methods that are diverse and efficient in computational time. After the selection of three base DL methods, a fuzzy rank-based logic using two non-linear functions is developed for the aggregation of ensemble outputs. The functions used for fuzzy ranking are of different concavities. Hence, both penalization and reward strategies are leveraged in the proposed fuzzy ensemble.

To develop the benchmark dataset, first, the reviews of cars were collected from Ganesan et al. [9]. Further, the co-reference resolution was applied to the review sentences to ensure that all the aspects discussed in the reviews are covered in the dataset. Further, the sentences were manually annotated based on guidelines released for benchmark datasets by SemEval [4]. An extensive evaluation of the proposed fuzzy rank-based ensemble approach is performed on five benchmark datasets including the newly proposed dataset of cars.

Intending to advance the research in the field of ALSC, the major contributions of this study are 1. Manual annotation of the car data for ALSC in the automobile domain. This release of data will push forward cross-domain transfer learning and automated ABSA research in the automobile domain. 2. A novel ensemble fuzzy-based approach using the three base deep learning (DL) methods which are diverse and efficient in terms computational time. Further, the systematic penalization and reward strategy ensures the prediction of the correct class by the proposed ensemble. 3. Experimental results demonstrate that the proposed ensemble approach performs significantly better than the base learners as well as other state-of-art deep learning methods. 4. This study also presents case studies to show that the proposed ensemble can predict correctly even when all the base learners give wrong predictions.

The organization of this study is as follows. Section 2 describes the related work. Section 3 describes the approach used to develop the dataset. Section 4 presents the hybrid fuzzy ensemble approach proposed in this study. Section 5 provides experimental details, statistical test results, and an ablation study. Section 6 discusses the case studies. Finally, section 7 concludes this work.

2 Related work

2.1. Deep learning for ALSC

In recent years, the literature on ALSC is primarily dominated by methods deploying deep neural networks. This dominance of deep learning methods is mainly because of their capability of learning features automatically without any external feature engineering effort [8] [9]. Additionally, such methods have shown remarkably better performance as compared to traditional machine learning methods [10]. The initial attempts in the deep learning based ALSC utilized sequence networks like LSTM in their architectures [11] [12]. Tang et al. [11] proposed the very first LSTM based model known as target-dependent LSTM(TD-LSTM). TD-LSTM can capture the context on both sides of the aspect or target in a sentence. Later, Wang et al. [12] proposed an attention-based model with LSTM as the underlying network in the architecture. The authors were inspired by the popularity of attention mechanism in the field of NLP and were the first to leverage attention mechanism in the area of ALSC. Followed by their work, various attention-based models using GRU, CNN, and memory networks [13] are proposed for ALSC task. Tang et al. [14]developed a network known as MemNet that utilizes a deep memory network to generate aspect-specific features and updated the memory using the attention mechanism. With the capability of capturing local features efficiently, Convolutional neural networks (CNN) based models have also demonstrated promising performance for ALSC task [10]. The researchers have leveraged simple CNN and CNN with hybrid architectures to achieve promising results. Xue et al. [15] proposed a method based on CNN and used the gating mechanism to efficiently handle the flow of information. Li et al. [16] proposed TNet, a hybrid architecture-based transformation network based on both CNN and LSTM. In TNet, LSTM is used to capture the contextual information whereas CNN captures the local features.

There are various other attempts in which interactive attention-based networks are proposed [17] [18] [19]. The intuition behind the interactive attention is to capture the relationship between the context and aspect of the sentence. The authors of [17] [18] argued that simple attention is not sufficient to capture the relationship between aspect and context. In another attempt, Fan et al. [19] proposed a coarse-grained and fine-grained attention mechanism referred to as a multigrain attention network. Recently, graph neural networks (GNN) have also gained importance for ALSC task. The researchers working in the area of ALSC, leverage GNNs to incorporate the syntactical knowledge of the sentence obtained from the dependency tree. Syntactic knowledge plays a crucial role in handling long-range dependencies between aspect and relevant context words. The graph-like structure of the dependency tree facilitates the usage of GNN for this task. However, the architecture of these GNN-based methods is quite complex which makes them computationally expensive. The very first attempts in this line are made by Zhang et al. [20] and Huang et al. [21]. The more recent works leveraging GNNs are [22] [23] [24] [25]. Initially, only node information of the sentence was captured using

GNN-based architectures. However, the edges of the dependency tree also carry important and meaningful information. Thus, in various works [26] [27], the edge information is also taken into consideration to generate a better representation of the sentence.

Table 1: Summary of the related work

Method	Year	Description	Average Acc[1]	Advantages	Limitation
TD-LSTM	2016	It has two LSTMs for handling the left and right context of the target	71.83	Simple architecture, less computational time	Low accuracy
AEContextAvg	2019	A simple feed-forward network that takes an average of aspect and context embeddings as input	74.36	Simple architecture, less computational time	Moderate accuracy
ATAE-LSTM	2016	Append the aspect embeddings with LSTM along with the attention layer.	70.66	Simple architecture, moderate computational time	Very low accuracy
CNN	2019	Simple network based on CNN to extract local features efficiently	70.80	Simple architecture, less computational time	Very low accuracy
MemNet	2016	Based on a memory network where context words act as external memory	71.02	Moderate computational time	Low accuracy
RAM	2017	Based on GRU and Bi-LSTM for generating aspect and context representation respectively	71.40	Moderate computational time	Low accuracy
IAN	2017	Based on two LSTMs along with interactive attention	72.33	Moderate computational time	Low accuracy
TNet	2018	Combines the embeddings generated by Bi-LSTM and CNN with a transformation module	75.39	Moderate accuracy	Complex architecture, high computational time
ASGCN	2019	Converts syntactic information into an undirected graph and then applies GCN	81.29	Very high accuracy	Complex architecture, high computational time
DualGCN	2022	Based on two GCNS: Syntactic and semantic	81.59	Very high accuracy	Complex architecture, high computational time
SSEGCN	2022	Generates the attention scores using two different types of attentions and further passes it to GCN layer	82.30	Very high accuracy	Complex architecture, high computational time
Ensemble majority vote	2022	Base learners: TD-LSTM, AEContext_Avg, and CNN, ensemble creation using majority voting method	72.30	Simple ensemble computation	Low accuracy
Stacking based ensemble	2022	Base learners: TD-LSTM, AEContext_Avg, and CNN, ensemble creation using the meta-learning approach with random forest classifier	76.74	Simple architecture of base learners	Moderate accuracy, Meta-learning increases the computational time
EO based ensemble	2022	Optimization approach applied to select base deep learner from a pool of ten DL methods, ensemble creation using meta-learning approach with random forest classifier	78.05	High accuracy	Complex computation for ensemble creation
Proposed Fuzzy Ensemble	2023	Base learners: TD-LSTM, AEContext_Avg, and CNN, Ensemble creation using simple mathematical fuzzy logic	79.61	Simple architecture of base learners, Simple logic for ensemble, High accuracy	Our proposed method will require future research on the scalability of our method across many more corpora of products and services.

[1] Very low: acc below 71; low: 71≤acc<74; moderate: 74≤acc<77; high:77≤acc<80; very high: acc above 80

2.2 Ensemble learning in the context of ALSC

Ensemble learning is a popular technique that has attracted researchers in most domains throughout the years. However, there are very few ensemble approaches proposed for ALSC task so far. Mohammadi et al. [30] were the first to propose an ensemble-based technique for the ALSC task. The authors used the simple CNN, Bi-LSTM, LSTM, and GRU as the base learners in their approach. Their ensemble approach was based on the meta-learning principle that fuses the prediction of the base learners to get the final prediction for the ensemble. However, their work has two major limitations. First, the authors selected simple deep neural networks and did not emphasize the aspect-specific information in any of the base learners. Second, the authors demonstrated the performance of the ensemble using the macro-precision metric only. However, accuracy and F1 score are considered better metrics to evaluate the performance of classification models.

Sharma et al. [28] proposed another meta-learning ensemble technique for ALSC. The authors used three base learners in their ensemble that are TD-LSTM, AEContextAvg, and CNN. Further, a random forest is used in the meta-learning phase to generate the final predictions. In another similar attempt, the authors of [29], proposed an ensemble approach based on the principle of classifier ensemble reduction. The authors transformed the selection of the base learner method for the ensemble as a classifier reduction problem. Further, a physics-based optimization algorithm known as the EO (Equilibrium Optimizer) algorithm is used to select the base learners from the pool of ten different DL methods. The EO-based ensemble obtained good performance as well. However, the selection of base learners using an optimization algorithm can be time taking and tedious process. Therefore, in this study, the aim is to propose an ensemble that is efficient as well less complex in terms of computation and time as well.

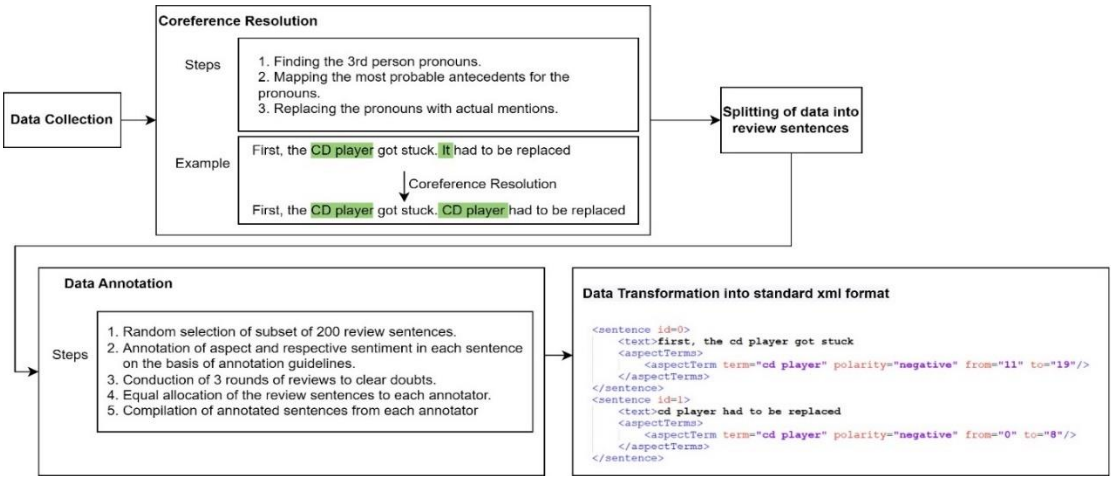
A brief description of various DL methods in ALSC literature along with their advantages and limitations is provided in Table 1. The computational time for various methods can be obtained from the work of Sharma & Kaur [31]. It can be seen from the table that most of the methods with simple architecture and less computational time could not achieve higher accuracy. At the same time, the methods attaining higher accuracies either have a complex architecture with high computational time or have complex ensemble construction methods.

Thus, in this work, the aim is to propose a method with improved accuracy and less complex computations. Therefore, three simple base learners are selected for the ensemble and further, a novel mathematical logic based on the fuzzy principle is proposed in this study.

3 Data annotation

3.1 Data collection

In this study, a new automobile domain is explored for ALSC research. The data collection and annotation process are carried out in a similar manner as to SemEval datasets. The sentences are annotated from the Car review dataset collected by Ganesan et al. [32]. The dataset contains the reviews of cars from the website named *‘caredmunds.com’*. The full reviews of one model for each car company are selected from the dataset. Further, before beginning with the annotation process, coreference resolution is applied to the reviews so that all the mentioned aspects in the reviews can be considered. Fig. 1 shows the elaborated steps followed for the data (construction) preprocessing task.



1. Finding the 3rd person pronouns.
2. Mapping the most probable antecedents for the pronouns.
3. Replacing the pronouns with actual mentions.

First, the CD player got stuck. It had to be replaced

First, the CD player got stuck. CD player had to be replaced

1. Random selection of subset of 200 review sentences.
2. Annotation of aspect and respective sentiment in each sentence on the basis of annotation guidelines.
3. Conduction of 3 rounds of reviews to clear doubts.
4. Equal allocation of the review sentences to each annotator.
5. Compilation of annotated sentences from each annotator

```
<sentence id=0>
    <text>first, the cd player got stuck
    <aspectTerms>
        <aspectTerm term="cd player" polarity="negative" from="11" to="19"/>
    </aspectTerms>
</sentence>
<sentence id=1>
    <text>cd player had to be replaced
    <aspectTerms>
        <aspectTerm term="cd player" polarity="negative" from="0" to="8"/>
    </aspectTerms>
</sentence>
```

Figure 1: Data Annotation Process

3.2 Co-reference resolution

In ALSC, the annotation is carried out at the sentence level. If the sentence contains any explicit target mentions, then the target is annotated. Otherwise, if the target mention is implicitly referred to using pronouns, then the sentence is not considered for annotation because of this, some valuable opinions may be missed. Co-reference resolution refers to the task of matching the expressions with the same entity in the text. This task helps in linking the actual noun phrase mention with its pronoun in the sentence. The co-reference resolution on the car reviews data collected from [32] is applied in this study. This step ensures that all the mentioned 3rd person pronouns like 'it', 'them', etc. are replaced with their actual mentions. The example showing the utility of co-reference resolution is shown in Fig.1. The task of co-reference resolution is performed using the NeuralCoref library available in Python. Further, the reviews are split into sentences and the data annotation task is performed.

3.3 Data annotation

In this step, each sentence is parsed to find the relevant aspect and its respective sentiment. For annotating the sentences, the annotation guidelines released by the organizers of the SemEval workshop [3] are followed. This ensures that the annotated data is at par with other benchmark datasets of ALSC task. The task of annotating the sentences is carried out with the help of two annotators who are the authors of this paper. Both annotators are initially required to annotate a subset of sentences based on the guidelines released by SemEval2. Then, a review session is conducted to discuss the annotations and other doubts. After 3 such rounds of discussions, all the doubts got clear to both annotators. Later, the rest of the review sentences are equally allotted to both annotators. Finally, the annotated review sentences are combined to form the final dataset. The sample annotated sentences from the dataset are shown in Table 2. After the annotation process is complete, the data is transformed into the standard XML format as in the benchmark datasets of ALSC literature. This step ensures that the prepared dataset can be easily used just like other benchmark datasets by researchers working in the area of ALSC. Random split is applied to get the train and test data respectively. The total number of sentences in the dataset is 5478 whereas the total number of samples is 7541.

Table 2: Sample sentences from the dataset

Review Sentence	Aspect Term	Polarity
Unfortunately, after rolling just 19k, transmission crapped out.	Transmission	-1
My only complaint is rear seats are not comfortable on back for a long car trip.	Rear seats	-1
Get the factory navigation system if you can.	Navigation system	0
Get back up camera for sure.	Back up camera	0
Fit and finish inside and out is fantastic.	Fit and finish	1
The interior is well layed out with easy to read gauges.	Interior	1

The statistics of the final released dataset are mentioned in Table 3.

Table 3: Data statistics

Car Dataset	Positive	Negative	Neutral	Total Samples	Number of sentences
Train	3253	1004	795	5052	4404
Test	1603	494	392	2489	1074

4 Methodology

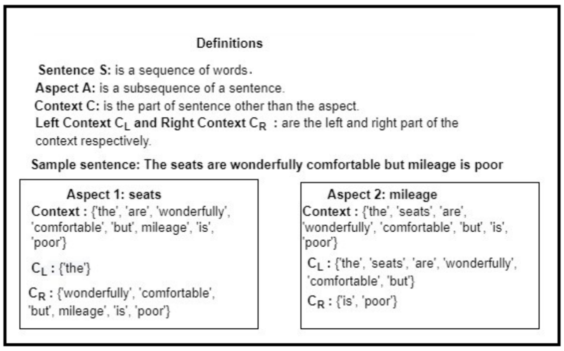
In this section, in the first step, the task definition along with the preliminaries related to ALSC and the deep learning methods is explained. In the second step, the different deep learning methods used as base classifiers in the proposed fuzzy ensemble approach are explained. Lastly, the proposed fuzzy ensemble approach based on selected DL methods is explained.

4.1 Preliminaries

The process of aspect sentiment classification is different from the general sentiment classification task. The major reason behind this difference is the presence of different polarity words for different aspects present in a single sentence. For example, in Fig. 2, *"The seats are wonderfully comfortable but the mileage is poor",* the sentiment polarity is positive for aspect *"seats"* and negative for aspect *"mileage"*. Thus, the ALSC task deals with predicting the polarity class for given pair of sentence

[2] The guidelines are available at: http://alt.qcri.org/semeval2014/task4/data/uploads/.

and aspect (S, A). In deep learning-based ALSC, the sentence S is converted into an output vector while taking aspect A into consideration. The construction of this output vector which is also the final representation of the input sentence varies depending on the underlying architecture of the deep learning method. Finally, this output vector is treated as the final feature and is fed into the softmax layer for sentiment prediction of the aspect A. Fig. 3 shows the overall architecture of the proposed fuzzy ensemble approach.



Definitions

Sentence S: is a sequence of words.

Aspect A: is a subsequence of a sentence.

Context C: is the part of sentence other than the aspect.

Left Context C_L and Right Context C_R : are the left and right part of the context respectively.

Sample sentence: The seats are wonderfully comfortable but mileage is poor

Aspect 1: seats

Context : {'the', 'are', 'wonderfully', 'comfortable', 'but', mileage', 'is', 'poor'}

C_L : {'the'}

C_R : {'wonderfully', 'comfortable', 'but', mileage', 'is', 'poor'}

Aspect 2: mileage

Context : {'the', 'seats', 'are', 'wonderfully', 'comfortable', 'but', 'is', 'poor'}

C_L : {'the', 'seats', 'are', 'wonderfully', 'comfortable', 'but'}

C_R : {'is', 'poor'}

Figure 2: Preliminaries for ALSC task

4.2 Selection of base deep learning methods

The performance of any ensemble method majorly depends on the selection of the base learners. Time efficiency plays a crucial role as the objective is to build an ensemble with better accuracy without compromising in terms of time. Further, to ensure that the selected base learners are diverse, different base learners and their characteristics discussed in [2] are thoroughly studied. After closely analysing the limitations in current approaches, three deep learning methods are selected that are diverse and time efficient. The three selected base learner methods are TD-LSTM, AEContextAvg, and CNN.

CNNs have the capability to extract local features efficiently. In most of the deep learning-based architectures of ALSC literature, the convolutional layer is placed after the input embedding layer to generate the local features from the text. Thus, CNN is chosen as one of the base learners in the proposed ensemble. Another base learner, TD-LSTM performs well for long sentences as it considers both the right context and left context of the aspect term to predict the sentiment polarity. To handle simple and short sentences, AEContextAvg is chosen. AEContextAvg has a very simple architecture and performs decently on short sentences.

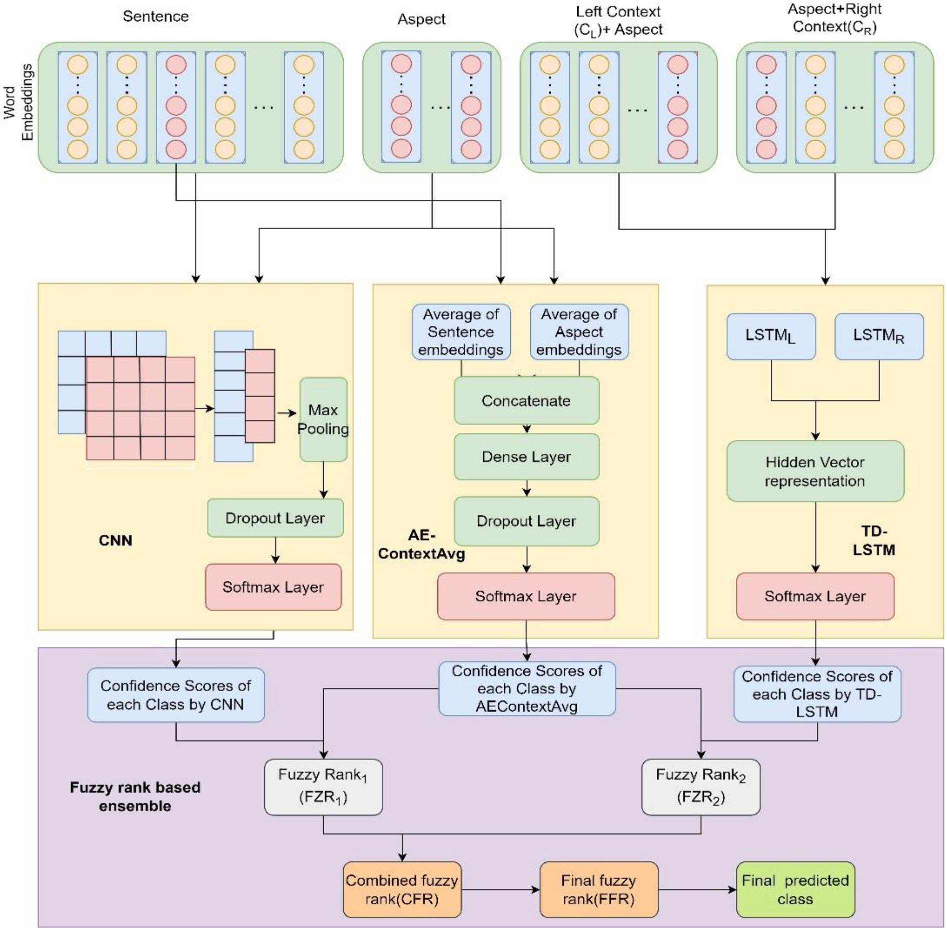

Sentence
Aspect
Left Context (C_L)+ Aspect
Aspect+Right Context(C_R)
Word Embeddings
Max Pooling
Dropout Layer
CNN
Softmax Layer
Average of Sentence embeddings
Average of Aspect embeddings
Concatenate
Dense Layer
Dropout Layer
AE-ContextAvg
Softmax Layer
$LSTM_L$
$LSTM_R$
Hidden Vector representation
TD-LSTM
Softmax Layer
Confidence Scores of each Class by CNN
Confidence Scores of each Class by AEContextAvg
Confidence Scores of each Class by TD-LSTM
Fuzzy rank based ensemble
Fuzzy $Rank_1$ (FZR_1)
Fuzzy $Rank_2$ (FZR_2)
Combined fuzzy rank(CFR)
Final fuzzy rank(FFR)
Final predicted class

Figure 3: The architecture of the proposed fuzzy-based ensemble.

These three methods are briefly explained below.

Target dependent LSTM

The TD-LSTM handles the input in an aspect-oriented way by splitting the input sentence around the aspect into the left context and right context. This way, TD-LSTM ensures considering the aspect position while generating the final feature vector of the sentence. As shown in Fig. 3, there are two separate inputs provided to 2 LSTMS, the LSTM$_L$ takes left context(C_L)+ aspect as input whereas LSTM$_R$ takes aspect+right context(C_R) as input. Finally, the hidden vector representation from both the LSTMs is combined to form the final vector further fed into softmax for prediction.

AEContextAvg

AEContextAvg [10] utilizes simple feed-forward neural network architecture which takes both aspect and sentence as input. As shown in Fig. 3, first, the average of aspect and sentence vectors are concatenated. Then this concatenated vector is fed into the softmax layer for prediction. The architecture of AEContextAvg is simple yet efficient because it considers both the aspect and sentence together.

Convolutional neural network

CNN is a deep learning network that acquires its power from the convolution filters and pooling operations. CNN is proven to be efficient in automatically extracting features from the text. as well. The first layer of CNN is an embedding layer that takes the word embeddings of the sentence as the input. Later, the output is fed into multiple convolution filters as shown in Fig. 3. Finally, the softmax layer generates the confidence score of each class for the given input.

4.3 Proposed fuzzy ensemble

In the proposed fuzzy ensemble, the confidence score (class probability) of the predictions of each base learner is considered and then this score value is utilized to calculate the final prediction for each sample.

A fuzzy rank is calculated for each class using two non-linear functions: $tanh$ and the modified Weibull function [33]. The chosen two functions in the proposed approach are of different concavities. The different concavities of the functions help in maintaining the equilibrium between the reward and penalization strategy. These functions determine the fuzzy rank of the various classes using the confidence scores obtained by each class for each base deep learning method.

Algorithm 1: Fuzzy Rank based Ensemble

Input: Probability scores for each class obtained by each base DL classifier

Output: Final Predicted Class

```
$p$ represents the number of classes and
$q$ represents the number of base DL
classifiers.
$i$ represents the of $i^{th}$ class and
$j$ represents the $j^{th}$ base DL classifier.
Initialize the $p\ X\ q$ list $PBS_j^i$ with the
confidence scores obtained for each class
$i$ by each base DL classifier $j$.
Initialize $FZR1_j^i$ and $FZR2_j^i$to store the
two fuzzy ranks obtained for each class $i$
by each base DL classifier $j$.
Initialize $CFR_j^i$ to store the combined
fuzzy score obtained using $FZR1_j^i$ and
$FZR2_j^i$
Initialize $FFR^i$ to store the final fuzzy
rank obtained by each class.
for each class $i$ and base DL classifier $j$
do
    Use Eq. (1) and (2) to calculate
    $FZR1_j^i$ and $FZR2_j^i$ respectively.
    Use Eq. (3) to calculate the
    combined fuzzy score $CFR_j^i$.
    Calculate the final fuzzy rank $FFR^i$
    using Eq. (4).
end for
return final predicted class =
$\min(FFR^i)$ for $i \in [1,p]$
```

The mathematical model used in this work is discussed next.

Let p represents the number of classes and q represents the number of base DL classifiers.

Initialize the $p\ X\ q$ list PBS_j^i to store the confidence score c_j^i obtained for the i^{th} class and j^{th} base DL classifier.

For each classifier, the two fuzzy ranks $FZR1_j^i$ and $FZR2_j^i$ are calculated using Eq. (1) and Eq. (2) based on the hyperbolic tangent and Weibull functions respectively.

$$FZR1_j^i = 1 - \tanh\left[\frac{(c_j^i - 1)^2}{2}\right] \quad \text{for } i \in [1,p] \qquad (1)$$

$$FZR2_j^i = \frac{\exp\left(-2\left(c_j^i\right)^2\right)}{2} \quad \text{for } i \in [1,p] \qquad (2)$$

Further, both the calculated fuzzy ranks are multiplied to obtain the combined fuzzy rank CFR_j^i as explained in Eq. (3).

$$CFR_j^i = FZR1_j^i \times FZR2_j^i \quad (3)$$

The final aggregated fuzzy rank for each class i is obtained using Eq. (4).

$FFR^i = \sum_{j=1}^{q} CFR_j^i$ for $j \in [1, \mathrm{q}]$ where q is the number of base DL classifiers. (4)

Finally, Eq. (5) is used to get the final predicted class by the ensemble.

Final predicted class= $\min(FFR^i)$ for $i \in [1, p]$ (5)

The detailed steps of the proposed fuzzy rank-based ensemble are explained in Algorithm 1.

5 Experiments

5.1 Experimental settings

The implementation of the proposed work is performed using the PyTorch framework. The hyperparameter details are shown in Table 4. The other model-specific parameter settings are kept the same as in [2]. Accuracy and Macro-F1 score are used as the evaluation metrics.

Table 4: Hyperparameter settings

GloVe embedding dimension	300
Hidden state vector dimension	300
Batch size	64
Learning Rate	0.01
Regularization	L2
Dropout rate	0.1
Optimizer	Adam
Initialization of weight matrix	U (-0.01,0.01)

Throughout this paper, acc refers to accuracy and F1 score refers to macro F1 score. The reliability of results is ensured by taking an average of 5 runs with randomly initialized values.

5.2 Datasets

In this study, the evaluation of various methods is performed on 5 datasets as mentioned in Table 5. The first four datasets: restaurant14, laptop14, restaurant15, and restaurant16 are released by SemEval whereas the Car dataset is developed in this study itself.

Table 5: Details of the datasets

Dataset	Positive		Negative		Neutral	
	Train	Test	Train	Test	Train	Test
Laptop 14	987	341	866	128	460	169
Restaurant 14	2164	728	805	196	633	196
Restaurant 15	1198	454	403	346	53	45
Restaurant 16	1657	611	749	204	101	44
Car	3253	1603	1004	494	795	392

5.3 Experimental results

In this section, the experimental results obtained by various methods are presented. The proposed ensemble approach is compared with various state-of-art baselines including the latest GNNs and ensemble methods of ALSC literature. It can be observed from Table 6 that the proposed fuzzy ensemble has outperformed most of the compared methods with few exceptions.

- The three base learners in the proposed model: TD_LSTM, CNN, and AEContextAvg have simple architecture without any attention mechanism or complex graph neural networks. Since TD_LSTM and CNN do not employ any attention mechanism, their performance is relatively poor as compared to other DL methods. However, the third base learner AEContextAvg has attained moderate accuracy even without the attention mechanism.

- The proposed fuzzy ensemble has outperformed all three base learners that are TD-LSTM, AEContextAvg, and CNN by 14.4 %, 13.7 %, and 14.2% respectively in terms of F1 score. This good performance of the proposed ensemble in comparison to base learners justifies the concept that weak and diverse base learners contribute to a good ensemble. Thus, the selection of base deep learning classifiers in this study is also justified.

- The other state of art methods like ATAE-LSTM, MemNet, IAN, and RAM deploy different types of attention mechanisms in their architecture. Nevertheless, their performance is almost similar (or slightly better) to the TD_LSTM and CNN. However, our proposed fuzzy logic-based ensemble model has attained better results as compared to the above methods even with weak base learners like TD_LSTM and CNN.

Table 6: Experimental results obtained for various methods

Methods	Restaurant14		Laptop14		Restaurant15		Restaurant16		Car		Average Acc	Average F1 score
	Acc	F1 score	Acc	F1 score	Acc	F1 score	Acc	F1 score	Acc	F1 score		
CNN [10]	73.75	60.3	61.75	53.06	64.3	40.43	77.18	47.54	77.02	71.91	70.8	54.65
AEContexTAvg [10]	70.71	56.99	63.79	55.71	82.7	51.2	78.57	51.3	76.01	70.41	74.35	57.12
TD-LSTM [11]	71.78	58.05	63.0	56.98	65.79	45.36	79.39	52.31	79.21	72.09	71.83	56.96
ATAE-LSTM [12]	70.98	54.61	63.0	52.52	66.15	42.74	74.24	52.77	78.92	71.02	70.66	54.73
MemNet [14]	70.35	56.77	61.59	50.94	63.9	43.3	78.92	49.21	80.32	72.48	71.02	54.54
IAN [18]	70.71	56.49	62.53	57.08	69.23	46.11	79.39	54.77	79.8	67.23	72.33	56.34
RAM [13]	70.26	55.94	60.5	52.61	68.16	45.42	79.04	50.2	79.46	72.01	71.48	55.24
TNet [16]	78.75	67.54	71	64.9	60.15	57.71	86.13	68.82	80.92	74.21	75.39	66.64
ASGCN [20]	81.69	73.76	75.02	70.79	78.96	60.71	87.71	67.83	83.07	76.32	81.29	69.88
DualGCN [22]	83.24	75.22	76.61	72.96	80.88	65.32	84.61	69.05	83.11	76.02	81.69	71.71
SSEGCN [23]	83.35	76.03	78.01	73.21	81.27	64.46	85.3	68.9	83.55	75.68	82.30	71.66
Ensemble majority vote	73.21	62.01	65.42	59.02	68.85	50.11	78.03	51.55	75.98	67.09	72.30	57.96
Stacking based ensemble [28]	81.07	76.32	70.23	68.11	73.3	70.71	80.08	61.4	79.01	72.35	76.74	69.78
EO based ensemble [29]	81.89	77.65	71.07	69.34	77.53	71.01	81.47	61.6	78.3	72.41	78.05	70.41
Proposed Fuzzy Ensemble	82.51	77.89	72.03	70.01	78.88	72.13	82.5	62.07	82.12	75.01	79.61	71.42

- As per the table, the performance of the proposed ensemble in terms of F1 score is better than the state-of-art methods like ATAE-LSTM, MemNet, IAN, and RAM by 16.4%, 16.6%, 14.8%, and 15.9% respectively.
- TNet method being a state-of-the-art method in ALSC literature has attained good performance. The architecture of TNet is quite complex whereas our proposed ensemble model has simple DL methods as base learners. Nevertheless, our proposed ensemble has outperformed TNet by 4.3%.
- The above compared methods did not incorporate the syntactic knowledge for ALSC task. Syntactic knowledge plays a crucial role in mapping correct opinion words to the aspect. Thus, the performance of methods without syntactic knowledge is quite less as compared to methods with syntactic knowledge like ASGCN, DualGCN, and SSEGCN. Graph neural networks are the most suitable networks for incorporating syntactic knowledge because the dependency tree of the sentence can be easily fed as a graph to such methods. Thus, ASGCN, DualGCN, and SSEGCN utilize various layers of GCN and have very complex architecture. Their performance is

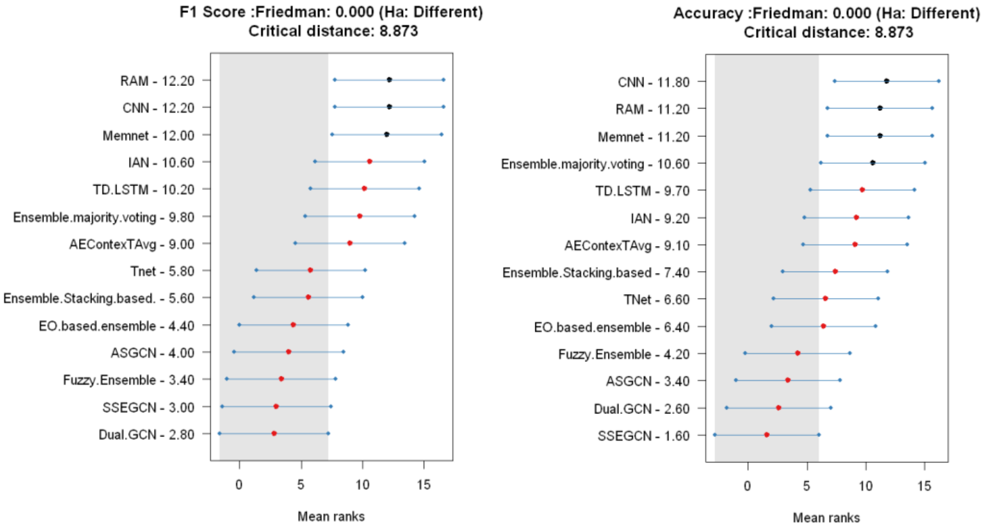
good but computational time is more. Even with simple base learners, our proposed ensemble model has reached comparable performance (if not better) with such GNN based methods.

- The proposed fuzzy ensemble is also compared with three other types of ensemble methods proposed in previous ALSC literature. The experimental results show that the proposed fuzzy ensemble approach has outperformed all other ensemble methods that are majority voting, stacking-based, and EO-based ensembles, with the same base learners.
- The proposed fuzzy ensemble has outperformed the simple majority voting ensemble with a difference of 13.2%. The majority voting ensemble directly works on the predicted classes and no emphasis is given to the confidence score obtained by each class. In contrast to this, our proposed fuzzy ensemble works minutely on the confidence scores thereby leading to better performance.
- The stacking-based ensemble is based on the confidence scores of predictions which are further combined using the principle of stacking. The random forest classifier is adopted in their work [28] to compute the final predicted class where the confidence scores obtained by base learners are considered features. This stacking or meta-learning-based approach increases the overall complexity of the ensemble. In contrast to this, our proposed ensemble is based on fuzzy logic where simple mathematical steps are followed to obtain the final predictions. Nevertheless, the performance of the latter is better with a difference of 1.5% and 2.9% for F1 score and accuracy respectively.
- The EO-based ensemble works on the principle of optimization using a heuristic approach where base learners are selected using the EO approach and later random forest classifier is applied in similar manner as in the stacking-based ensemble approach. Therefore, the overall complexity of this ensemble is even more than the stacking-based ensemble. In contrast to this, our proposed fuzzy ensemble is simple yet efficient and has clearly outperformed the EO-based ensemble.
- It can be said that a common issue with ensemble learning models is high computational complexity. This work aims to propose a simple yet computationally efficient ensemble aggregation approach. Our proposed ensemble is simple yet efficient.
- The phenomenon of better performance shown by the proposed ensemble also demonstrates that the ranking strategy in the proposed ensemble is quite efficient in predicting the correct class, even when all the base learners give wrong predictions.

For a better understanding of this phenomenon, two case studies are presented in section 6.

5.4 Statistical test

In this study, statistical testing is performed to validate the proposed ensemble approach. The statistical test applied is the Friedman test which is a non-parametric counterpart of ANOVA. The Nemenyi test is used as the post-hoc test for the Friedman test. The Nemenyi test is conducted after the rejection of the null hypothesis of the Friedman test.



a) Nemenyi Test for F1 Score

b) Nemenyi Test for Accuracy

Figure 4: Post-hoc test results

The null and alternate hypotheses of the Friedman test are:

Hypothesis 1 (H1). The performance of the compared methods is not significantly different in terms of F1 score and accuracy.

Alternate Hypothesis(H1a): At least two of the compared methods have significant differences in F1 score and accuracy.

The statistical tests are performed using the R tool. The average ranks of various methods across all datasets are used to conduct the Friedman test. As per the Friedman test results, the null hypothesis is rejected for both evaluation metrics: F1 score and accuracy. Therefore, the post-hoc Nemenyi test is conducted to find the methods that have a statistically significant difference in their performance. Fig. 4(a) and 4(b) show the results of the post-hoc test.

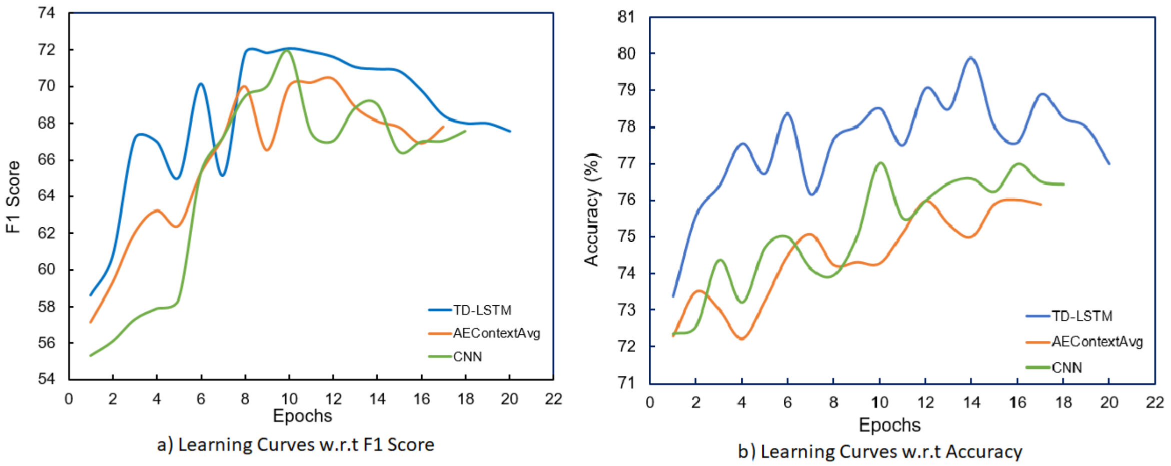
In Fig. 4(a) and 4(b), the compared methods are plotted against the average rank obtained by each method. DualGCN and SSEGCN are the best-ranked methods for F1 score and accuracy respectively. The average rank of the proposed fuzzy ensemble is 3.4 and 4.2 for the F1 score and accuracy respectively. The methods with the lines that fall within the grey area indicate that the performance difference is not statistically significant. Therefore, it can be concluded that even though the proposed fuzzy ensemble could not outperform the DualGCN or SSEGCN methods, it is no worse than DualGCN, SSEGCN, and ASGCN methods based on the Friedman test. Thus, the proposed ensemble based on simple deep learning classifiers has achieved comparable performance with top-performing complex GNN based methods.

5.5 Ablation study

Ablation experiments were carried out to demonstrate that the proposed work performs better than either the ensemble of any of the two base classifiers or the performance of individual classifiers. The three classifiers were combined in every conceivable way for this study. The accuracy and F1-score for each combination are calculated using the proposed ensemble model for all five datasets as shown in Table 7.

Table 7: Ablation experiment results

Methods	Restaurant14		Laptop14		Restaurant15		Restaurant16		Car	
	Acc	F1 score	Acc	F1 score	Acc	F1 score	Acc	F1 score	Acc	F1 score
CNN	73.75	60.3	61.75	53.06	64.3	40.43	77.18	47.54	77.02	71.91
AEContextAvg	70.71	56.99	63.79	55.71	82.7	51.2	78.57	51.3	76.01	70.41
TD-LSTM	71.78	58.05	63.01	56.98	65.79	45.36	79.39	52.31	79.21	72.09
CNN+TD-LSTM	75.43	73.31	67.58	64.02	69.04	67.34	80.82	58.87	79.85	72.69
CNN+AEContextAvg	76.61	73.06	66.51	65.23	71.32	68.09	79.07	58.32	79.47	73.24
TD-LSTM+AEContextAvg	77.02	74.89	68.34	64.51	72.46	67.65	80.55	59.03	80.01	73.03
Fuzzy ensemble (CNN+TD-LSTM+AEContextAvg)	**82.51**	**77.89**	**72.03**	**70.01**	**78.88**	**72.13**	**82.5**	**62.07**	**82.12**	**74.01**



a) Learning Curves w.r.t F1 Score

b) Learning Curves w.r.t Accuracy

Figure 5: Learning Curves of base deep learning methods

It can be easily inferred from the results that applying the proposed ensemble logic to combine the three basic classifiers produces the best results out of all potential combinations, supporting the rationale for their selection.

5.6 Learning curves

The learning curves of the three base deep learning methods for the car dataset are presented in this section. Fig. 5(a) and 5(b) show the learning curves for F1 score and accuracy respectively. Early stopping is applied for training all three base deep learning methods where the training is stopped once model performance stops improving on the validation set. As per Fig. 5(a) and 5(b), TD-LSTM, AEContextAvg, and CNN took 20, 17, and 18 epochs respectively to converge. It can be observed from Fig 5(a) and 5(b) that the performance of TD-LSTM in terms of accuracy is better than both CNN and AEContextAvg whereas the performance of all three methods in terms of F1 score is quite competitive. It can also be observed that TD-LSTM, AEContextAvg, and CNN obtained the best results at epochs 11, 12, and 10 respectively.

6 Case study

The intuition behind the fuzzy ensemble approach is to give weightage to the confidence score attained by each class irrespective of the final prediction. This way, the aim is correctly predicting the class even if all three base learners failed to do so. The usage of non-linear functions of different concavities helps in restrictively penalizing the classes. For a comprehensive understanding of the proposed approach, two case studies are presented in this section. The ranks FZR_1 and FZR_2 are calculated for each class using Eq. (1) and Eq. (2) (as defined in section 4.3) respectively. FZR_1 rewards for a high confidence score whereas FZR_2 penalize for the same. Next, CFR (refer to Eq. (3) of section 4.3) is calculated using FZR_1 and FZR_2. Finally, CFR obtained for each class by each base classifier is summated to get the FFR score (refer to Eq. (4) of section 4.3) where the class with a minimum value of FFR is considered as the predicted class.

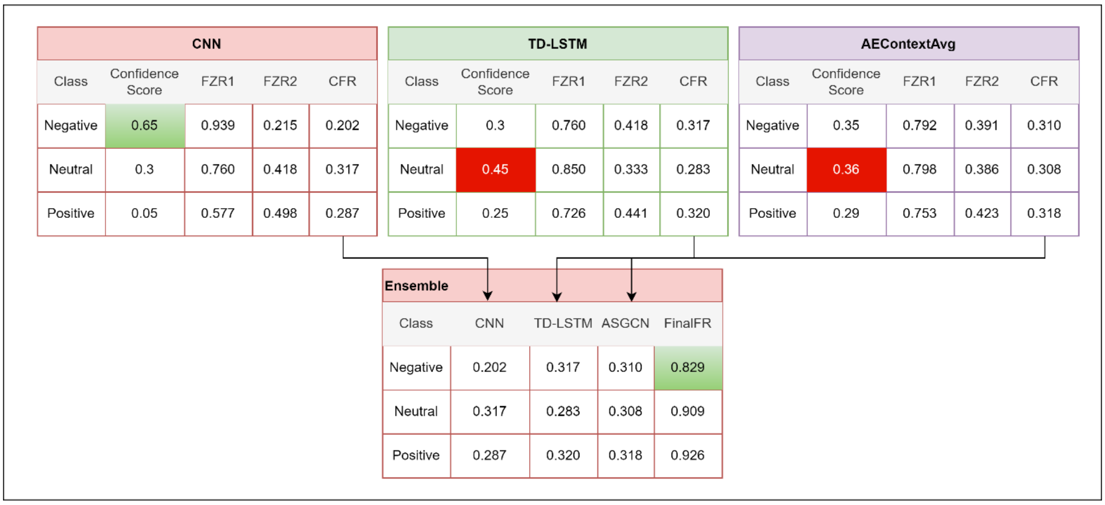
Case study 1: One base learner predicts correct class and two base learners predicted incorrect class

In the example shown in Fig. 6, one base learner predicts correctly with a moderate confidence score whereas two other classifiers predict incorrectly with less confidence score. In such a scenario, the proposed approach facilitates the prediction of the correct class on the basis of the confidence score attained by the correct class by each base learner, whereas a majority voting ensemble will fail. Fig. 6 shows the step-by-step calculation of the prediction made by the proposed approach.

In the above example, CNN predicted the correct class with a confidence score of 65 percent. However, the other two methods predicted another class 2 with the average confidence score ranging from 35-45 percent. Further, the two methods have given weights in the range of 30-35 percent to class 1 which is the correct prediction.

The proposed approach successfully predicted the correct class using the principle of restrictive penalization and rewarding class 1 on the basis of its confidence score obtained by each base learner. The above example shows the utility of the proposed approach when the majority base learner predicts the wrong class and the correct class losses with a very less margin.

Similarly, the proposed approach also has the potential to correct the misclassified samples even when all the base learners fail to predict correctly as discussed in case study 2.



CNN

Class	Confidence Score	FZR1	FZR2	CFR
Negative	0.65	0.939	0.215	0.202
Neutral	0.3	0.760	0.418	0.317
Positive	0.05	0.577	0.498	0.287

TD-LSTM

Class	Confidence Score	FZR1	FZR2	CFR
Negative	0.3	0.760	0.418	0.317
Neutral	0.45	0.850	0.333	0.283
Positive	0.25	0.726	0.441	0.320

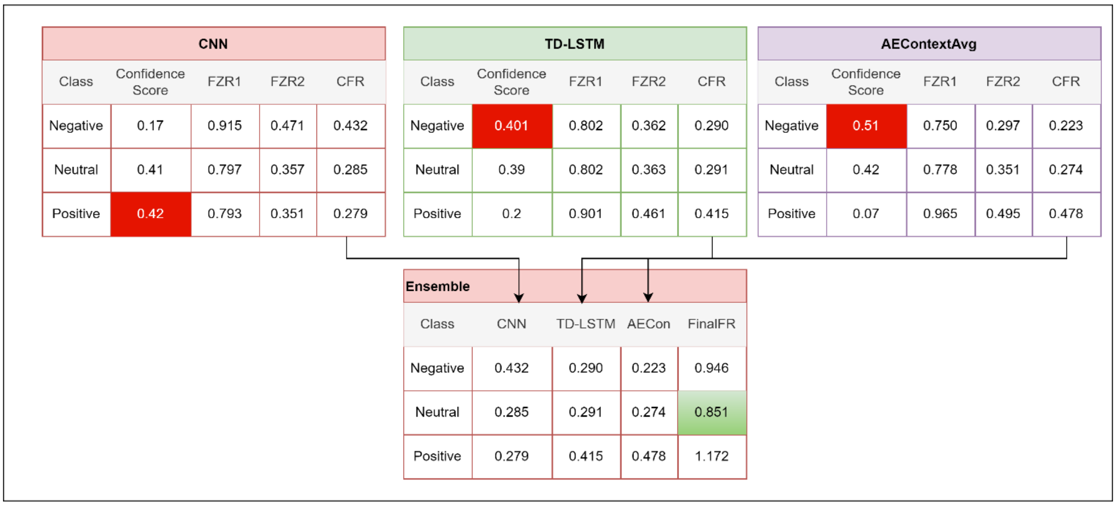
AEContextAvg

Class	Confidence Score	FZR1	FZR2	CFR
Negative	0.35	0.792	0.391	0.310
Neutral	0.36	0.798	0.386	0.308
Positive	0.29	0.753	0.423	0.318

Ensemble

Class	CNN	TD-LSTM	ASGCN	FinalFR
Negative	0.202	0.317	0.310	0.829
Neutral	0.317	0.283	0.308	0.909
Positive	0.287	0.320	0.318	0.926

Figure 6: Case study 1



CNN

Class	Confidence Score	FZR1	FZR2	CFR
Negative	0.17	0.915	0.471	0.432
Neutral	0.41	0.797	0.357	0.285
Positive	0.42	0.793	0.351	0.279

TD-LSTM

Class	Confidence Score	FZR1	FZR2	CFR
Negative	0.401	0.802	0.362	0.290
Neutral	0.39	0.802	0.363	0.291
Positive	0.2	0.901	0.461	0.415

AEContextAvg

Class	Confidence Score	FZR1	FZR2	CFR
Negative	0.51	0.750	0.297	0.223
Neutral	0.42	0.778	0.351	0.274
Positive	0.07	0.965	0.495	0.478

Ensemble

Class	CNN	TD-LSTM	AECon	FinalFR
Negative	0.432	0.290	0.223	0.946
Neutral	0.285	0.291	0.274	0.851
Positive	0.279	0.415	0.478	1.172

Figure 7: Case study 2

Case study 2: All three base learners predict incorrectly.

In this case, the correct class 2 has obtained a confidence score in the range of 39-45 percent. However, it fails to get the first rank by each base learner by a very slight margin. The calculations of the proposed approach in Fig. 7 show that the rewarding strategy of the approach using $tanh$ function (Eq. (1) in section 4.3) helps class 2 get the final score better than the other two classes.

7 Conclusions

In this study, a novel fuzzy-based ensemble technique is proposed for ALSC task. In addition to this, the Car dataset of the automobile domain is also developed as per the SemEval guidelines of benchmark datasets. The following are the conclusions of this study: A car dataset is developed in the automobile domain to facilitate automated and cross-domain learning in ALSC literature. The size of the proposed car dataset is around 7500 samples which is larger than other benchmark datasets available for ALSC task. Since the training of neural networks requires large datasets, this car data will also facilitate better training and testing of various deep learning methods proposed for ALSC task.

A novel fuzzy-based ensemble is proposed which utilizes the confidence scores of the classes predicted by each base learner. The proposed fuzzy ensemble is based on mathematical logic where two functions of different concavities are used to calculate the final predicted class. Further, the case studies presented in section 6 have validated the significance of the restrictive rewarding and penalization strategy used in this work. The performance of the proposed fuzzy ensemble technique is either better or at par with other top-performing latest deep learning-based methods.

In the future, fine-tuning of the hyperparameters of base learners can be performed to get better results. It is also suggested to incorporate more advanced deep neural architectures in the ensemble.

References

[1] K. Schouten and F. Frasincar, "Survey on Aspect-Level Sentiment Analysis," IEEE Transactions on Knowledge & Data Engineering , vol. 28, no. 3, pp. 813-830, March 2016. https://doi.org/10.1109/tkde.2015.2485209

[2] T. Sharma and K. Kaur, "Benchmarking Deep Learning Methods for Aspect Level Sentiment Classification," Applied Sciences, vol. 11(22):10542, November 2021. https://doi.org/10.3390/app112210542

[3] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," in 8th International Workshop on Semantic Evaluation (SemEval 2014), Dublin, Ireland, 2014. https://doi.org/10.3115/v1/s14-2004

[4] M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar and I. Androutsopoulos, "Semeval-2015 task 12: Aspect based sentiment analysis.," in 9th International Workshop on Semantic Evaluation (SemEval 2015), 2015. https://doi.org/10.18653/v1/s15-2082

[5] M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar and A. S. Mohammad, "SemEval-2016 task 5: Aspect based sentiment analysis.," in 10th international workshop on semantic evaluation (SemEval-2016), 2016. https://doi.org/10.18653/v1/S16-1002

[6] Q. Jiang, L. Chen, R. Xu, X. Ao and M. Yang, "A Challenge Dataset and Effective Models for Aspect-Based Sentiment Analysis," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019. https://doi.org/10.18653/v1/d19-1654

[7] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou and K. Xu, "Adaptive Recursive Neural Networkfor target-dependent twitter sentiment classification," in Proceedings of the 52nd annual meeting of the association for computational linguistics, Baltimore, Maryland, USA, 2014. https://doi.org/10.3115/v1/p14-2009

[8] W. Etaiwi, D. Suleiman and A. Awajan, "Deep Learning Based Techniques for Sentiment Analysis: A Survey," Informatica, vol. 45, no. 7, pp. 89-95, 2021. https://doi.org/10.31449/inf.v45i7.3674

[9] S. Al-Otaibi and A. Al-Rasheed, "A Review and Comparative Analysis of Sentiment Analysis Techniques," Informatica, vol. 46, no. 6, pp. 33-44, 2022. https://doi.org/10.31449/inf.v46i6.3991

[10] J. Zhou, J. X. Huang, Q. Chen, Q. V. Hu, T. Wang and L. He, "Deep Learning for Aspect-Level Sentiment Classification: Survey, Vision and Challenges.," IEEE Access, vol. 7, 2019. https://doi.org/10.1109/access.2019.2920075

[11] D. Tang, B. Qin, X. Feng and T. Liu, "Effective LSTMs for Target-Dependent Sentiment Classification," in Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, 2016.

[12] Y. Wang, M. Huang, L. Zhao and X. Zhu, "Attention-based LSTM for Aspect-level Sentiment Classification," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016. https://doi.org/10.18653/v1/d16-1058

[13] P. Chen, L. Bing, Z. Sun and W. Yang, "Recurrent Attention Network on Memory for Aspect Sentiment Analysis," in Conference on Empirical Methods in Natural Language Processing, 2017. https://doi.org/10.18653/v1/d17-1047

[14] D. Tang, B. Qin and T. Liu, "Aspect Level Sentiment Classification with Deep Memory Network," in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, 2016. https://doi.org/10.18653/v1/d16-1021

[15] W. Xue and T. Li, "Aspect Based Sentiment Analysis with Gated Convolutional Networks," in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018. https://doi.org/10.18653/v1/p18-1234

[16] X. Li, L. Bing, W. Lam and B. Shi, "Transformation networks for target-oriented sentiment classification," in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 2018. https://doi.org/10.18653/v1/p18-1087

[17] B. Huang, Y. Ou and K. M. Carley, "Aspect Level Sentiment Classification with Attention-over-Attention Neural Networks," in International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation,SBP-BRiMS 2018, 2018. https://doi.org/10.1007/978-3-319-93372-6_22

[18] D. Ma, S. Li, X. Zhang and H. Wang, "Interactive Attention Networks for Aspect-Level Sentiment

Classification," in IJCAI'17: Proceedings of the 26th International Joint Conference on Artificial Intelligence,2017. https://doi.org/10.48550/arXiv.1709.00893

[19] F. Fan, Y. Feng and D. Zhao, "Multi-grained attention network for aspect-level sentiment classification," in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 2018. https://doi.org/10.18653/v1/d18-1380

[20] C. Zhang, Q. Li and D. Song, "Aspect-based Sentiment Classification with Aspect-specific Graph Convolutional Networks," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, 2019. https://doi.org/10.18653/v1/d19-1464

[21] B. Huang and K. M. Carley, "Syntax-Aware Aspect Level Sentiment Classification with Graph Attention Networks," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019. https://doi.org/10.18653/v1/d19-1549

[22] R. Li, H. Chen, F. Feng, Z. Ma, X. Wang and E. Hovy, "Dual Graph Convolutional Networks for Aspect-based Sentiment Analysis," in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Online, 2021. https://doi.org/10.18653/v1/2021.acl-long.494

[23] Z. Zhang, Z. Zhou and Y. Wang, "SSEGCN: Syntactic and Semantic Enhanced Graph Convolutional Network for Aspect-based Sentiment Analysis," in Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, United States, 2022. https://doi.org/10.18653/v1/2022.naacl-main.362

[24] H. Wu, C. Huang and S. Deng, "Improving aspect-based sentiment analysis with Knowledge-aware Dependency Graph Network," Information Fusion, vol. 92, pp. 289-299, 2022. https://doi.org/10.1016/j.inffus.2022.12.004

[25] B. Liang, H. Su, L. Gui, E. Cambria and R. Xu, "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," Knowledge-Based Systems, vol. 235, p. 107643, 2022. https://doi.org/10.1016/j.knosys.2021.107643

[26] X. Bai, P. Liu and Y. Zhang, "Investigating Typed Syntactic Dependencies for Targeted Sentiment Classification Using Graph Attention Neural Network," IEEE/ACM Transactions on Audio, Speech, and Language Processing, pp. 503-514, 2021. https://doi.org/10.1109/taslp.2020.3042009

[27] X. Zhu, L. Zhu, J. Guo, S. Liang and S. Dietze, "GL-GCN: Global and Local Dependency Guided Graph Convolutional Networks for aspect-based sentiment classification," Expert Systems With Applications, vol. 186, 2021. https://doi.org/10.1016/j.eswa.2021.115712

[28] T. Sharma and K. Kaur, "An Ensemble approach for Aspect level sentiment classification using deep learning methods," in 3rd International Conference on Data Analytics & Management (ICDAM-2022), 2022. https://doi.org/10.1007/978-981-19-7615-5_69

[29] T. Sharma and K. Kaur, "An Equilibrium Optimizer based Ensemble for Aspect level Sentiment Classification," in presented at International Conference on Advances and Applications of Artificial Intelligence and Machine Learning(ICAAAIML), 2022.

[30] A. Mohammadi and A. Shaverizade, "Ensemble Deep Learning for Aspect-based Sentiment Analysis," International Journal of Nonlinear Analysis and Applications, vol. 12, no. Special Issue, Winter and Spring 2021, pp. 29-38, 2021. https://doi.org/10.22075/IJNAA.2021.4769

[31] T. Sharma and K. Kaur, "Aspect sentiment classification using syntactic neighbour based attention network," Journal of King Saud University - Computer and Information Sciences, vol. 35, no. 2, pp. 612-625, 2023. https://doi.org/10.1016/j.jksuci.2023.01.005

[32] K. Ganesan and C. Zhai, "Opinion-based entity ranking," Information Retrieval, vol. 15, no. 2, pp. 116-150, 2012. https://doi.org/10.1007/s10791-011-9174-8

[33] H. Rinne, The Weibull Distribution A Handbook, 1st ed., Chapman & Hall, 2020.