Received: December 8, 2021. Revised: February 23, 2022.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Polarity Detection and Feature-Based Summarization for Aspect Based Sentiment Analysis Dataset

Hetal Vipul Gandhi[1,2]* **Vahida Zakirhusen Attar[1]**

[1]*Department of Computer Engineering and IT, College of Engineering, Savitribai Phule Pune University, Pune, India*
[2]*Department of Computer Science and Engineering, Walchand College of Engineering, Sangli, India*
* Corresponding author's Email: hetalvgandhi@rediffmail.com

Abstract: Aspect Based Sentiment Analysis (ABSA) aims to perceive the opinion of the majority of reviewers for each feature (aspect category) of the product. We target both Aspect Term Polarity (ATP) and Aspect Category Polarity (ACP) subtasks of ABSA. We retrieve different word vector-based representations of sentence and augment them with n-gram and sentiment features using novel algorithm. We employ supervised machine learning approaches and propose the ensemble of Logistic Regression, k-Nearest Neighbor, and Support Vector Machine models for these subtasks. We experiment with four major domains from Hindi ABSA Dataset and report increase in performance accuracy by 16% for ATP subtask and 2% to 13% across domains for ACP subtask over results from researchers who released this dataset. Secondly, we also propose a novel WFBOS method to compute the Summary Score [range: -100 to 100] of each category from a particular domain. We make use of frequency of positive, neutral, negative, and conflict sentiment, of each aspect category from among the reviews of a particular domain in Hindi ABSA Dataset to do this computation.

Keywords: Sentence vector, Aspect category, Feature-based summarization, Sentiment analysis, Ensemble model.

1. Introduction

On many e-commerce and social networking websites, stakeholders exchange their feedback or thoughts. When it comes to analysing the feedback or reviews for any product and summarizing it, the document-based [1] or sentence-based [2, 3] sentiment analysis was performed on the set of reviews. It is coarse-grained analysis and focuses on determining the sentiment of the reviewer for a document or sentence in terms of *positive*, *negative* or *neutral*. However, such analysis allows us to associate the sentiment to the product and not to the feature(s) of the product. To accomplish such a goal, aspect based sentiment analysis (ABSA) need to be performed. Previously, it was also known as feature-based opinion summarization [4]. The features of the product may be explicit- *aspect term* or implicit- *aspect category*. Initially, the six primary features defined for *Restaurants* domain were called '*Aspect*' instead of '*Aspect category*' [5]. They made use of support vector machine models for sentiment classification. The correlation of classifier output classes and metadata in the form of star rating was computed to determine the quality of the models.

The ABSA task [6] involves four subtasks: two related to aspect terms and two related to aspect categories. The dataset released had annotations for English reviews for *Laptops* and *Restaurants* domains. The sample annotated review from restaurants domain from [6] is as shown in Fig. 1. As can be seen from Fig. 1, we have addressed the task of polarity detection for both aspect terms and aspect categories for Hindi ABSA dataset [7, 8]. The set of polarity levels include the *conflict* polarity level in addition to basic three polarity levels- *neutral*, *positive*, and *negative*.

The lexicon-based approach is mostly used for polarity related subtasks [9]. The lexicon created in [10] was used for sentiment classification [11]. The

Received: December 8, 2021. Revised: February 23, 2022.

```xml
<sentence id="2200">
    <text>The design and atmosphere is just as good.</text>
    <aspectTerms>
        <aspectTerm term="design" polarity="positive" from="4" to="10"/>
        <aspectTerm term="atmosphere" polarity="positive" from="15" to="25"/>
    </aspectTerms>
    <aspectCategories>
        <aspectCategory category="ambience" polarity="positive"/>
    </aspectCategories>
```

Figure. 1 Sample annotated review from *Restaurant's* domain of SemEval 2014 ABSA dataset

lexicon-based approach is suited if domain specific lexicon is available. It is not the case for ABSA subtasks for Hindi ABSA dataset which contains reviews from 12 different domains [7] forming four major domains of ABSA category dataset [8].

Bilgin and Koktas [12] experimented with the five machine learning based models using word2vec and doc2vec based features. Among different machine classifiers, mostly SVM classifiers are found to be effective for such tasks [13]. Few other models used for sentiment analysis tasks are maximum entropy classifiers [14], KNN classifiers [15], [16], Naïve bayes [7]. The ensemble of different classifiers and different feature sets was also proposed [2] for the sentiment analysis task. The machine learning models provided improved performance when traditional features were combined with word vector-based features when analyzing English reviews for the described literature. We augment such representations with novel additional features for targeted subtasks.

For reviews in non-English languages like Dutch, French, Russian etc., the deep learning based-approach was used for similar ABSA task [17]. They made use of convolutional neural network followed by pooling and SoftMax activation. For Arabic [18], and Czech [14] reviews, the sentiment classification was performed using the different machine learning models. For Hindi, the ABSA dataset is released and the sentiment classification task is attained using machine learning models [7], [8]. The SVM model provided accuracy of 54.05 % for ST2 of ABSA [7]. For ST4 of ABSA, the accuracy reported ranges from minimum 47.95 % to maximum 91.62 % across four domains, using decision tree based classifier. The results are good enough for the *Movies* domain for ST4 (91.62 %), but for the other domains of ST4 and for ST2, there is further scope for improvement. We have targeted to improve this performance accuracy for both these subtasks.

The novel method of predicting the normalized score of product sentiment was used for recommendation [19]. Hu and Liu [20] were the first to talk about generating summary for each feature of the product after doing feature-wise analysis. For this purpose, they targeted to represent the group of reviews which are *positive* or *negative* with their count under separate heads for each feature, along with the link to those actual reviews. It can be useful if there is wide variation among these frequency values of different aspect categories of the product. However, to get better insights, the method of deriving text rating by using the frequencies of *positive* and *negative* sentences was proposed [5]. This text rating value was compared with the value obtained through other regression methods.

To generate a summary from multiple documents containing evaluative text, Carenini proposed two methods based on sentence extractor and natural language generation. Two set of features including crude and user-defined features were used. The clustering approach was used to derive summary score [21, 22]. Kansal and Toshniwal [21] applied this approach to classify reviews into two classes for each feature and further derive star rating. On the other hand, the similar approach was used to derive rating on scale of [0, 5] for each aspect [22]. Such rating was ultimately used to derive product rating by considering count of reviewers referencing each feature and assigning weights accordingly. As there is no summarization method which considers frequency values of all four polarity levels to derive summary score, we propose a method for computing Summary Score of each Domain-category pair by assigning weights to each polarity level.

We also put forth the different approaches used for polarity related Hindi ABSA subtasks to improve the performance accuracy. We target two subtasks- *Aspect Term Polarity* and *Aspect Category Polarity*. The major contributions of the paper include:

- We propose an ensemble model: a linear combination of logistic regression (LoR), support vector machine (SVM), and K-nearest neighbor (KNN) models for these subtasks.
- We experiment with different sentence vector representations and other novel features to build

Received: December 8, 2021. Revised: February 23, 2022.

the models.

- We propose *SBF Algorithm* to retrieve sentiment-based features using Hindi SentiWordNet.
- We also compare the results of Term polarity with Transfer Learning based ULMFiT models [23].
- We propose a Weighed Feature-based opinion summarization (WFBOS) method that accounts for frequency values of all four polarity levels to derive summary score in the range of -100 to 100, for each aspect category of a particular domain in this paper.

In section 2, we describe the proposed methodology for targeted subtasks after feature extraction methods, using the different machine learning approaches. In section 3, we share the results of ABSA subtasks for Hindi reviews and compare them with state-of-the-art results. In section 4, we conclude and discuss future enhancements.

2. Proposed methodology

ABSA includes four subtasks- two of them related to *aspect term* and two of them related to *aspect category*. The basic kind of problem addressed for polarity related subtasks is the multi-class classification problem with four sentiment classes. We discuss the features extracted and proposed methodology for polarity related ABSA subtasks. We also share the mathematical model to derive a summary score (SSoC) for each aspect category from particular domain, known the associated or predicted polarity of reviews.

2.1 Feature extraction

The features extracted to build the model for the targeted subtasks for Hindi reviews include-

- *Word vectors*: Retrieved for Hindi Vocabulary words from FasText[1]
- *Sentence Vector*
- *Sentiment-based Feature*
- *N-gram Features*
- *Category Association Word Features*

The sentence vector representation is derived using word embeddings of all the words in the review. We compute the sentence vector of the sentence *s*, having *l* words, where for each j^{th} word, the word vector is denoted by v^j and v_i^j represents i^{th} component of v^j. We experiment with different word vector based representation but based on performance select the sum of vector-based representation as given in Eq. (1).

$$SV\_Sum(s) = \Big\|_{i=1}^{d} \sum_{j=1}^{l} v_i^j \qquad (1)$$

We use N-gram features and sentiment features for polarity detection subtasks. We extract unigram and bigram features and employ feature selection method to restrict the features to minimum count of 100 using the chi-squared feature selection method. The important characteristic of such n-gram feature extraction method is that it returns n-gram feature counts from training and test data over the same set of features, after applying this feature selection over training data alone.

For aspect category polarity subtask, the category association word features are also derived by measuring the cosine similarity of category word with each word in the review sentence from training data. As category words are in English, they are translated into the corresponding Hindi word or set of words. For each Hindi category word(s), the cosine similarity value is computed and sorted. From this set, the features having the maximum similarity score are considered.

Boumhidi and Nfaoui [9] made use of a lexicon for classifying tweets at entity-level into three classes. They also made use of part-of-speech of words for this task. In a similar manner, to determine the polarity of aspect category, we also propose the SBF algorithm (Algorithm 1) to retrieve the sentiment-based feature for each review sentence.

We use Hindi SentiWordNet [24], a lexicon which contains the *positive* and *negative* polarity score mentioned along with its part-of-speech for Hindi words from the corpus. The *PoS_tag_list* is assigned to restrict to only adjective, verb, and adverb as part-of-speech of words. We also identify seven words which are prone to appear in the review having *conflict* polarity like 'लेकिन', 'पर', 'हालांकि' etc. and assign it to *con_pol_List.* The Get_Sentiment_Features() method is called with these two variables as parameters, in addition to SentiWordNet resource handle and review sentence for which sentiment feature is to be retrieved.

First, the entries from resource are received for all words in the *Dict_List* variable. If review contains any word from the *con_pol_List*, the conflict polarity is directly assigned to the review sentence. Otherwise, for each word in the review

[1] https://fasttext.cc/docs/en/pretrained-vectors.html

Algorithm. 1 SBF algorithm- for sentiment-based features from Hindi SentiWordNet

```
Get_Sentiment_Features (Sent, Hin-
Input: @Sent- Review sentence
@Hindi SentiWordNet - An online resource
@Con_pol_list- List of words appearing in
@PoS_Tag_list- List of allowed part-of-speech of
Output: Sentiment Feature using value associated
Get list of words Word_List from @Sent
Get list of entries Dict_List from @Hindi
p_pol = 0.0 and p_cnt = 0
n_pol = 0.0 and n_cnt = 0
For each w in Word_List
    If w in @Con_pol_List
        Pol = 'con'
    else
        Tag, pos, neg = Dict_List(w)    //
        // Assert Tag is in @PoS_Tag_list
        If pos > neg and pos!=0.0
            p_pol += pos
            p_cnt += 1
        Else If neg > pos and neg!=0.0
            n_pol += neg
            n_cnt += 1.0
If p_cnt> n_cnt and p_pol > n_pol
    Pol = 'pos'
else if n_cnt> p_cnt and n_pol > p_pol
    Pol = 'neg'
Else Pol = 'neu'
Return (Pol)
```

sentence, we retrieve its positive and negative polarity score from *Dict_List*, if its part-of-speech belongs to *PoS_tag_list*. We maintain the count of positive and negative polarity words (*p_cnt*, *n_cnt*) in the review and update the sum of positive and negative polarity score (*p_pol*, *n_pol*), using scores of words appearing the review sentence from *Dict_List*. The final polarity assigned is as given in Eq. (2). This polarity pol value is returned by the Get_Sentiment_Features() method. It is finally also used as one of the features for Category polarity detection subtask of ABSA.

$$pol = \begin{cases} 'pos', \ if\ p\_cnt > n\_cnt\ and\ p\_pol > n\_pol \\ 'neg', \ if\ n\_cnt > p\_cnt\ and\ n\_pol > p\_pol \\ 'neu' \quad otherwise \end{cases} \tag{2}$$

2.2 Aspect term polarity models

As a pre-processing, we discard reviews having no aspect terms. We evaluate the performance in two ways- 70-20-10 for training-testing-validation as in [25], [26] and 3-fold cross validation as in [7], [8]. We use 70-10-20 split when working with ULMFiT models. The universal language model fine-tuning (ULMFiT) and MultiFit models are proposed for addressing the aspect term polarity subtask [23]. It involves a sort of transfer learning-based technique to get better model parameters. ULMFiT model includes building the pre-trained model and fine-tuning it for the aspect term polarity classification task [27].

The primary steps involved in such ULMFiT models are- discriminative fine-tuning, gradual unfreezing and slanted triangular learning rates. It makes use of three-layer LSTM architecture- AWD-LSTM [28]. We first create a language model for Hindi using large corpus from Wikipedia. We then fine-tune it for our Hindi ABSA dataset [7] in the second step. Finally, we train our classification model [23] using encoder weights of our fine-tuned model.

We also split the dataset into 3-folds and build the model using ensemble of LoR, SVM and KNN models for this subtask. The probability values returned from each classifier-pr_i for aspect term-*atm*, referenced in sentence *s*, are used to compute the sentiment associated with *atm*, using Ensemble as given in Eq. (3). Here, pr_1, pr_2 and pr_3 denote the probabilities returned by LoR, SVM and KNN models for having term polarity tp_i. The polarity having maximum probability sum, is assigned to aspect term. Assume *TP* denotes the set of polarity levels. Finally, the sentiment polarity assigned for the aspect term *atm* in the sentence *s* is the polarity tp_i for which $sent(s, atm, tp_k) = 1$.

$$sent(s, atm, tp_k) = \begin{cases} 1, \ if \ \max\limits_{tp_i \in TP} \begin{bmatrix} pr_1(tp_i|s, atm) \\ +pr_2(tp_i|s, atm) \\ +pr_3(tp_i|s, atm) \end{bmatrix} occurs\ for\ i = k \\ 0, \ otherwise \end{cases} \tag{3}$$

2.3 Aspect category polarity models

The aspect category polarity subtask targets classifying the polarity of an aspect category into one of the four polarity classes. The different features used to build the classification model are character and word unigrams & bigrams, their tf-idf values, sentence vector (from word embeddings), sentiment score, and category association word-based features. The steps involved for extracting unigram and bigram features from the train and test set with feature selection are explained in section 2.1. As features associated with aspect category are also required, we extract category association word-based features as discussed in section 2.1.

The aspect category class labels also contribute to the feature space when determining polarity of the Aspect category. The following machine learning models with their Ensemble are created using 3-fold cross-validation- LoR, SVM and KNN. The polarity assigned to the review sentence is one with the maximum sum of the probabilities from three classifiers for the needed (category, polarity) pair. The polarity label assignment for sentence *s* and category *c* with Ensemble is given in Eq. (4).

$$Senti\_Label(s,c) = \begin{cases} pol, if\ p_1(pol|s,c) + p_2(pol|s,c) + p_3(pol|s,c) \\ \quad is\ maximum\ for\ pol\ \in Polarity\_Set \\ \quad 0,\ otherwise \end{cases} \tag{4}$$

The term $p_1(pol|s,c)$ denotes the probability obtained from LoR model for polarity *pol* for the sentence-category pair *(s-c)*. The other two classifiers in sequence are SVM and KNN. By using probability values returned by three classifiers, the majority voting method was also used for this subtask.

2.4 Feature-based summarization

In this subsection, we discuss a method for feature-based opinion summarization after analysing the set of reviews from a particular domain. Usually, the frequency of the polarity of aspect categories is used to generate this summary. The simple representation of feature-based opinion summary using frequency of four polarity levels is as shown in Fig. 2. We have used values from an *Electronics* domain of sample Hindi ABSA dataset [8].

Instead of visualizing this summary in the form of frequency of each polarity level for each aspect category as in Fig. 2 or as described by Hu and Liu [20], we propose a novel method to compute the Summary score (SS), for each aspect category in range of -100 to 100 by using weights for each polarity level. Such WFBOS method to compute summary score is useful in following ways-

- To compare the *SS* of different aspect categories of a particular product
- To compare the *SS* of same aspect category for different brands of the same product

To derive this *SS* value, we need frequency of polarity levels for a particular aspect category. Let us say- *FPos*, *FNeg*, *FCon*, and *FNeu* denote the frequency of *positive*, *negative*, *conflict*, and *neutral* polarities for aspect category c_i. We assign -100 weight to *negative* polarity and 100 weight to *positive* polarity. We assume that the weight for *neutral* polarity is *WNeu* instead of being 0. We propose *Proportion based Method* to derive *WNeu.* The proportion of positive aspect categories over sum of *positive* and *negative* aspect categories is used to compute *WNeu* using Eq. (5).

$$WNeu = \left[\frac{FPos}{FPos+FNeg} \times 200\right] - 100 \tag{5}$$

We then use all the frequency values to compute the *BaseSum* which is indicative of the number of reviews contributing for summary score computation. The *BaseSum* computation is as given in Eq. (6). Here, conflict polarity samples contribute to twice its frequency value as they equally contribute to positive and negative polarity. As

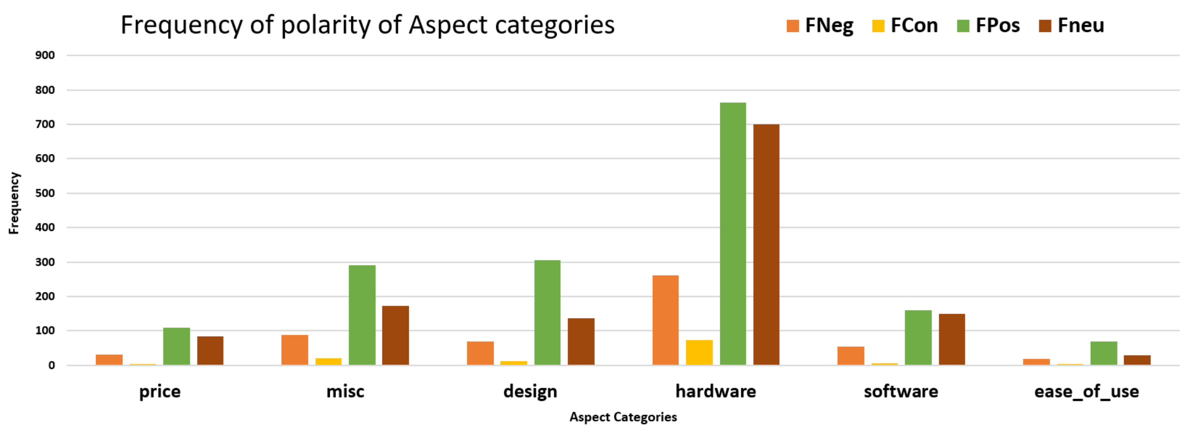

Figure. 2 Opinion summarization of the set of reviews from electronics domain of Hindi ABSA dataset

Received: December 8, 2021. Revised: February 23, 2022.

neutral samples may be considered to be either more *positive* or *negative*, they are directly added for *BaseSum* computation.

$$BaseSum = FNeg + 2 \times FCon + FPos + FNeu \quad (6)$$

The final summary score computation is targeted using weighed summation method and is as given in Eq. (7). The number of conflict reviews, *FCon* is not neglected as it has both negative and positive sentiment attached to it.

$$SS(cat) = \frac{\begin{bmatrix} -100 \times (FNeg + FCon) + 100 \times (FCon + FPos) \\ + WNeu \times FNeu \end{bmatrix}}{BaseSum} \quad (7)$$

3. Results and discussion

This section describes the Hindi ABSA datasets [7], [8], and the accuracy results obtained for the targeted subtasks. We also present the summary scores computed for *Electronics* domain of Hindi ABSA aspect category dataset.

3.1 Hindi ABSA dataset

The annotations provided for aspect terms [7] and aspect categories [8] constitutes the ABSA dataset. These subsets contain 5417 annotated Hindi review sentences. The reviews are from 12 domains provided with the respective annotations in the XML format. For the purpose of performance comparison of category based results, the nine domains having the same set of aspect categories are grouped and called *Electronics* domain [8]. The reviews from other three domains- *Mobile_apps*, *Travels*, and *Movies* are as is. The sample snapshot of a review sentence from term and category dataset is as shown in Fig. 3. *Accuracy* is used as a performance measure for polarity-based subtask evaluation.

```
<sentence id="cam_83">
    <text>फुजीफिल्म मिनी 50एस सेल्फी टाइमर के साथ आया है।</text>
    <aspectTerms>
        <aspectTerm from="20" polarity="pos" term="सेल्फी टाइमर" to="32"/>
    </aspectTerms>
```

(a)

```
<sentence id="cam_83" polarity="pos">
    <text>फुजीफिल्म मिनी 50एस सेल्फी टाइमर के साथ आया है।</text>
    <aspectCategories>
        <aspectCategory category="hardware" polarity="pos"/>
    </aspectCategories>
</sentence>
```

(b)

Figure. 3 Snapshot of annotated reviews from Hindi ABSA (a) Term dataset (b) Category dataset

Table 1. Accuracy score of aspect term polarity subtask on aspect term dataset with 3-fold cross-validation.

Folds	LoR	SVM	KNN	Ensemble
0	69.71	64.58	61.78	69.57
1	70.11	66.11	63.45	70.37
2	69.37	66.31	62.98	70.24
Average	69.73	65.67	62.74	**70.06**

Table 2. Domain-wise accuracy of ATP subtask using ensemble models

Domain Names	Akhtar et al. [7]	ULMFiT Models [23]	Ensemble Model
Laptops	50.98	**80**	7.55
Mobiles	54.07	**83.08**	69.5
Tablets	57.19	**80.26**	8.1
Cameras	57.06	**91.17**	7.22
Headphones	46.15	**85.71**	57.14
Home_ Appliances	79.23	0.6	**80.95**
Speakers	53.84	**90**	51.11
Smart_ Watches	64.7	**89.19**	15.02
Televisions	65.47	**90.48**	12.77
Mobile_apps	**61.53**	60	57.86
Travels	68.78	74.45	**78.81**
Movies	39.23	**60.23**	52.99

Table 3. Performance comparison of aspect term polarity subtask with state-of-the-art results using 3-fold cross-validation on Hindi aspect term dataset

Reference	Approach	% Accuracy
Akhtar et al. [7]	SVM	54.05
Proposed	Ensemble	70.06

3.2 Aspect term polarity results

The results obtained for LoR, SVM and KNN and their ensemble for this subtask are as given in Table 1. We employ 3-fold cross-validation to build the models and test them in each case.

We also put forth the results for 12 domains with 3-fold cross-validation for ATP subtask by training models from the same domain for Ensemble models. It can be observed that very few domains had improved performance for ensemble models. However, with ULMFiT models [23], where the model was trained with 70 % data, we report better performance for most of the domains as shown in Table 2.

 DOI: 10.22266/ijies2022.0831.02

The comparison of the results for aspect term dataset for aspect term polarity is as shown in Table 3. The results are derived using 3-fold cross-validation, the same splitting method which was used by Akhtar for Hindi ABSA term dataset in [7]. We report an accuracy of 70.06 % using ensemble models which is 16 % more than that of Akhtar [7].

3.3 Aspect category polarity results

We use a sentence vector derived using Eq. (1) by utilizing FastText word embeddings. We also extend this feature set by deriving n-gram and tf-idf features and applying the feature selection to it. As category specific polarity is to be determined, we also retrieve words closely associated with each aspect category and use them as features. For extracting sentiment-based feature using Hindi SentiWordNet, we use Algorithm 1. The performance accuracy obtained for aspect category polarity subtask using different models is compared in Table 4. We also split the whole dataset with stratified folds and then build models. The performance accuracy obtained for same is 66.15 %.

To get an intuition of feature contribution, we use ablation method. As can be seen in Table 5, there is maximum reduction in the accuracy for all the domains when sentence vector features are eliminated. Thus, it has its maximum contribution for the improved performance. The n-gram features do not contribute much as the dataset is smaller (for domains *Mobile_apps*, *Travels* and *Movies*).

3.4 Feature based summarization results

In order to grasp the overall reviewer sentiments associated with each aspect category of a product, we have proposed summary score computation as a WFBOS method in section 2.4. We defined the summary score computation method based on it. The summary score computation for all aspect categories from *Electronics* domain using Eqs. (4), (5) and (6) is as given in Table 6. Here, *FNeg*, *FCon*, *FPos* and *FNeu* represent the number of *negative*, *conflict*, *positive* and *neutral* categories for *Electronics* domain. Comparing summary score values from table, we can say that the *design* aspect category has a score of 60.05 which is *positive* and is maximum among all other aspect categories. Similarly, we can compare summary scores of aspect categories for any domain. The maximum summary score can be 100 and minimum can be -100 for any aspect category. For *Hardware* category, the score is *positive*, but it is minimum among score of other aspect categories, and so there is scope for improvement.

3.5 Discussion

From the result comparison of accuracy for term polarity subtasks, it can be observed that our proposed Ensemble and ULMFiT models provide considerable improvement over results in [7]. For the category polarity detection subtask, our ensemble model provides improved accuracy for all three domains except movies domain [8]. The sentence vector-based features maximally contribute for this improved performance. Our proposed summary score computation method is simple and is of great help to the community to enable comparison of performance of different categories of the same product. It can also be utilized for comparing the summary score of the same aspect category for different brands of similar products.

4. Conclusion and future work

This research work has focused upon improving the performance of polarity-based subtasks of Hindi ABSA, using Ensemble of machine learning models. From our study, we conclude that for such a dataset, the single machine learning model is not that efficient across all domains. So, we have proposed ensemble of three classifiers for these subtasks. The performance of feed forward neural network models is not as good as ensemble models across all domains. Among the contextual, syntactic and word embedding features used to build different models, it was observed that the word embedding based features are mostly responsible for improved performance. Our proposed WFBOS method for summary score computation method produces

Table 4. Comparison of accuracy of ensemble model and its classifiers with SOTA results for category polarity subtask

Domain	**LoR**	**SVM**	**KNN**	**Ensemble Model**	**Akhtar et al.** [8]
Electronics	65.53	62.46	57.83	**67.3**	54.48
Mobile_apps	50.77	49.23	48.21	**49.74**	47.95
Travels	68.08	62.79	61.55	**67.55**	65.2
Movies	76.35	72.64	72.75	76.13	**91.62**

Table 5. Average accuracy of 3-fold cross-validation for ACP subtask using ensemble model with ablation method

Elimination of	Electronics	Mobile_apps	Travels	Movies
SV	**64.51**	**49.74**	**64.02**	**74.89**
AW	66.83	50.77	66.84	76.24
N-Gram	65.59	54.36	70.02	76.35
SWN	66.83	51.28	66.49	76.46

Table 6. Summary score computation for electronics domain of Hindi ABSA category dataset [6] with *WNeu* set using proportion based method (maximum and minimum summary score of aspect category highlighted)

Category	FNeg	FCon	FPos	FNeu	WNeu	BaseSum	Summary Score
price	31	4	110	83	56.03	232	54.1
misc	89	21	290	173	53.03	594	49.28
design	69	13	305	137	63.1	537	**60.05**
hardware	261	73	763	700	49.02	1870	**45.2**
software	55	6	160	149	48.84	376	47.28
ease_of_use	19	3	70	30	57.3	125	54.55

values over a range -100 to 100. Such method can be used to compare this score of different aspect categories of the same product as well as the same aspect category of the same product of different brands. As a future work, we look out for employing deep neural networks like convolutional neural network and recurrent neural networks for polarity based subtasks.

Conflict of interest

The authors declare no conflict of interest.

References

[1] A. Joshi, A. R. Balamurali, and P. Bhattacharyya, "A Fall-back Strategy for Sentiment Analysis in Hindi: a Case Study", In: *Proc. of the 8th International Conference on Natural Language Processing*, 2010.

[2] O. Araque, I. C. Platas, J. F. S. Rada, and C. A. Iglesias, "Enhancing deep learning sentiment analysis with ensemble techniques in social applications", *International Journal Expert Systems With Applications*, Vol. 77, pp. 236–246, 2017.

[3] R. Abdullah, R. Sarno, and C. Fatichah, "Aspect-Based Sentiment Analysis for Sentence Types with Implicit Aspect and Explicit Opinion in Restaurant Review Using Grammatical Rules, Hybrid Approach, and SentiCircle", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 5, pp. 177-187, 2021, doi: 10.22266/ijies2021.1031.17.

[4] M. Hu and B. Liu, "Mining Opinion Features in Customer Reviews", In: *Proc. of The 19th National Conference on Artificial Intelligence - (AAAI'04)*, 2004, pp. 755–760.

[5] G. Ganu, N. Elhadad, and A. Marian, "Beyond the Stars : Improving Rating Predictions using Review Text Content", In: *Proc. of Twelfth International Workshop on the Web and Databases (WebDB 2009) J*, Vol. 9, No. WebDB, pp. 1–6, 2009.

[6] M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis", In: *Proc. of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, 2015, No. SemEval, pp. 27–35.

[7] M. S. Akhtar, A. Ekbal, and P. Bhattacharyya, "Aspect based Sentiment Analysis in Hindi : Resource Creation and Evaluation", In: *Proc. of the Tenth International Conference on Language Resources and Evaluation (LREC' 16)*, 2016, pp. 2703–2709.

[8] S. Akhtar, A. Ekbal, and P. Bhattacharyya, "Aspect Based Sentiment Analysis : Category Detection and Sentiment Classification for Hindi", In: *Proc. of International Conference on Intelligent Text Processing and Computational Linguistics*, 2016, pp. 246–257.

[9] A. Boumhidi and E. H. Nfaoui, "Leveraging Lexicon-Based and Sentiment Analysis Techniques for Online Reputation Generation", *International Journal of Intelligent Engineering and Systems*, Vol. 14, No. 6, pp. 274-289, 2021, doi: 10.22266/ijies2021.1231.25.

[10] T. Wilson, J. Wiebe, and P. Hoffmann,

"Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis", In: *Proc. of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, 2005, pp. 347–354.

[11] J. Yu, Z. J. Zha, M. Wang, and T. S. Chua, "Aspect ranking: Identifying important product aspects from online consumer reviews", In: *Proc. of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp. 1496–1505, 2011.

[12] M. Bilgin and H. Köktaş, "Sentiment analysis with term weighting and word vectors", *International Arab Journal of Information Technology*, Vol. 16, No. 5, pp. 953–959, 2019.

[13] A. A. Kumar, S. Kohail, A. A. Kumar, A. Ekbal, and C. Biemann, "IIT-TUDA at SemEval-2016 task 5: Beyond Sentiment lexicon: Combining domain dependency and distributional semantics features for aspect based sentiment analysis", In: *Proc. of 10th International Workshop on Semantic Evaluation, Proceedings*, pp. 1129–1135, 2016.

[14] T. Hercig, T. Brychc´ın, L. Svoboda, M. Konkol, and J. Steinberger, "Unsupervised Methods to Improve Aspect-Based Sentiment Analysis in Czech", *Computación y Sistemas,* Vol. 20, No. 3, pp. 365–375, 2016.

[15] S. Kaur, G. Sikka, and L. K. Awasthi, "Sentiment Analysis Approach Based on N-gram and KNN Classifier", In: *Proc. of the 1st International Conference on Secure Cyber Computing and Communications*, pp. 13–16, 2018.

[16] M. Rezwanul, A. Ali, and A. Rahman, "Sentiment Analysis on Twitter Data using KNN and SVM", *International Journal of Advanced Computer Science and Applications*, Vol. 8, No. 6, pp. 19–25, 2017.

[17] S. Ruder, P. Ghaffari, and J. G. Breslin, "INSIGHT-1 at SemEval-2016 Task 5: Deep learning for multilingual aspect-based sentiment analysis", In: *Proc. of 10th International Workshop on Semantic Evaluation, Proceedings*, pp. 330–336, 2016, doi: 10.18653/v1/s16-1053.

[18] O. Al-harbi, "Classifying Sentiment of Dialectal Arabic Reviews : A Semi-Supervised Approach", *International Arab Journal of Information Technology,* Vol. 16, No. 6, 2019.

[19] M. Syamala, "A Filter Based Improved Decision Tree Sentiment Classification Model for Real- Time Amazon Product Review Data", *International Journal of Intelligent Engineering and Systems*, Vol. 13, No. 1, pp. 191-202, 2020, doi: 10.22266/ijies2020.0229.18.

[20] M. Hu, B. Liu, and S. M. Street, "Mining and Summarizing Customer Reviews", In: *Proc. of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 168-177, 2004.

[21] H. Kansal and D. Toshniwal, "Aspect based summarization of context dependent opinion words", *Procedia Computer Science*, Vol. 35, No. C, pp. 166–175, 2014.

[22] Y. Lu, C. X. Zhai, and N. Sundaresan, "Rated aspect summarization of short comments", In: *Proc. of the 18th International Conference on World Wide Web*, pp. 131-140. 2009.

[23] H. V Gandhi and V. Z. Attar, "Transfer Learning for Aspect Term Polarity Determination", *Solid State Technology*, Vol. 63, No. 2, pp. 956–968, 2020.

[24] A. Das and S. Bandyopadhyay, "SentiWordNet for Indian Languages", In: *Proc. of the Eighth Workshop on Asian Language Resources*, pp. 56–63, 2010.

[25] M. S. Akhtar, A. Kumar, A. Ekbal, and P. Bhattacharya, "A Hybrid Deep Learning Architecture for Sentiment Analysis", In: *Proc. of COLING 2016*, pp. 482–493, 2016.

[26] M. S. Akhtar, P. Sawant, S. Sen, A. Ekbal, and P. Bhattacharyya, "Solving Data Sparsity for Aspect Based Sentiment Analysis Using Cross-Linguality and Multi-Linguality", In: *Proc. of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 572–582, 2018.

[27] J. Howard and S. Ruder, "Universal Language Model Fine-tuning for Text Classification", In: *arXiv preprint arXiv:1801.06146*, 2018.

[28] S. Merity, N. S. Keskar, and R. Socher, "Regularizing and Optimizing LSTM Language Models", In: *arXiv preprint arXiv:1708.02182*, 2017.