Journal of Theoretical and Applied Information Technology
30th June 2023. Vol.101. No 12
© 2023 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ASPECT-ORIENTED SUGGESTION MINING FROM OPINION REVIEWS

NAVEEN KUMAR LASKARI[1], SURESH KUMAR SANAMPUDI[2]

[1]Data Scientist, Xelpmoc Design and Tech, Research Scholar, JNTUH Hyderabad, Telangana, India

[2]Assistant Professor & HoD-IT, JNTUHCEJ, Nachupally, Jagtial

[1]naveen.laskari@gmail.com, [2]sureshsanampudi@gmail.com

ABSTRACT

Suggestion Mining refers to extracting suggestions from the opinionated text. The majority of existing research in suggestion mining does the binary classification of the given opinion review text as a suggestion or non-suggestion. The pioneering work ignored the fine-grained analysis, such as identifying the target towards which aspect the suggestion was mentioned or the target audience towards whom it intended to. However, such fine-grained analysis is more important in various use cases. A novel end-to-end hybrid model for fine-grained analysis of suggestions with aspect orientation has been proposed in this paper. We have utilized two different datasets of SemEval-2019 Task 9. The performance has been evaluated using other models from machine learning, neural network, and transfer learning with combinations of word embeddings. The experiment results demonstrate that our approaches worked well for aspect-oriented suggestion mining.

Keywords: *Suggestion Mining, Word Embeddings, Aspect Extraction, Text Classification, Rule-Based System*

1. INTRODUCTION

Opinion reviews are essential in service utilization, product purchase, and decision-making. Therefore, opinion reviews on social platforms became a reliable resource for diverse knowledge [12]. The benefits of opinion reviews on social platforms are multi-fold. It also helps fellow customers, third-party agents, and service providers [13]. It has been observed that different people's viewpoints include suggestions, hints, and recommendations on various matters. Suggestion Mining is identifying and extracting suggestions from customer reviews [19]. In the realm of Natural Language Processing (NLP), suggestion mining is a pretty modern subject. It aims to understand the nature of reviews, such as suggestions or non-suggestions, rather than evaluating user expression's positive or negative nature [20].

In the case of opinion reviews, the pioneering research focused on classifying the review as positive or negative at varying levels of granularity [2][30], Whereas, in suggestion mining, the current research efforts focus on classifying the opinion reviews as a suggestion or non-suggestion but not at the aspect level. Consider the example review from the travel domain "Then you would need a ***car or taxi*** to get to ***Graceland*** and other sights." This suggests having a car or taxi to visit a place, where the aspects are car or taxi and place is Graceland. This suggestion helps travelers plan their trip to a new place much better way. Consider another example *"You must present your **passport,** pay the number of **Euros**, and men must wear a **tie and jacket**"*. The review has suggestions towards various aspects such as ***passport, tie, and Jacket*** and ***Euros.*** *Similarly, consider the following reviews from the different domains, which have more aspects. "I can often get economical **flights** to **Hawaii or Mexico** without much notice from **Seattle**". "This **website** might give you some idea of the **cost of living** in **Turkey**". "You must be **careful** about your **things** and **money** while on **travel**".* Identifying these aspects expressed in the opinion review and the associated suggestions create a much higher business across the domains.

The models proposed in the literature classify these kinds of reviews as a suggestion but do not identify the abovementioned aspects. A recent survey report on suggestion mining revealed that rule-based, statistical, and partitioning-based models from machine learning, deep learning, pre-trained, and transfer learning methods are popularly applied. However, rule-based and machine learning-based approaches require handcrafted feature construction, which is labor-intensive [24]. Manual feature engineering does not apply to deep learning models, resulting in considerable success across the tasks in NLP, including suggestion mining. Furthermore, literature has witnessed that incorporating syntactic, linguistic, and stylistic features into deep learning models boosts the performance of various NLP tasks [3]. Identifying aspects across the domains and associated suggestions in opinion reviews to create enhanced business value motivated me to research suggestion mining and extract more insights. In researching suggestion mining, the following objectives are considered.

- ✓ Identify the most relevant aspects of the suggestions expressed in the opinion review.
- ✓ Methodology to deal with the class imbalance of suggestion mining.
- ✓ Reformulate the suggestion mining problem in relevance to aspect orientation.
- ✓ Build state-of-the-art models suitable for aspect-oriented suggestion mining.

The significant contributions of the paper to achieve the mentioned objectives are

- ✓ An ensemble of methods for aspect extraction from review using dictionary-based and unsupervised approaches.
- ✓ Handling class imbalance with the help of the over-sampling approach.
- ✓ Reformulation of suggestion mining task as a multi-label classification problem.
- ✓ Adopting machine learning and neural networks with pre-trained and transfer learning models to approach aspect-oriented suggestion mining.

The remaining parts of the paper are organized as follows; the systematic literature on suggestion mining is extended in section 2. A suitable methodology for the problem and algorithms are discussed in section 3. Section 4 presents the evaluation mechanism. Finally, section 5 reports conclusions and scope for future enhancement.

2. LITERATURE REVIEW

Sentiment Analysis and Opinion Mining is one of the most critical tasks in language processing applications. It is characterized as the examination of viewpoints, feelings, assessments, attitudes, appraisals, affects, perspectives, emotions, and subjectivity regarding various topics communicated in the text on various platforms. The pioneering work in sentiment analysis was carried out by [5] and [8]. In sentiment analysis, the major contributions went into classifying the sentiment at different levels, such as document level, sentence level, subjectivity level, and aspect level. In document-level classification, the entire document is assigned either a positive or negative label, regardless of the mixed opinions expressed in the entire document. At the sentence level, the sentiment is derived by considering the sentence as the primary unit; in Aspect Based Sentiment Analysis [4][5][10], the sentiments expressed in the text are evaluated towards a particular aspect or subject. In Aspect-Based Sentiment Analysis [5], the important phases of operations include aspect extraction, aspect category detection, aspect term polarity detection, and aspect category polarity detection [10]. This paper focused on aspect term extraction and identifying the suggestions for a specific aspect.

2.1 Aspect Term Extraction

The research community has extensively studied target extraction from opinion reviews as part of Aspect Based Sentiment Analysis (ABSA) and its applications. The literature mentions it as an aspect term extraction [16][21]. It applies all

Journal of Theoretical and Applied Information Technology
30th June 2023. Vol.101. No 12
© 2023 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

possible models beginning with a rule-based mechanism to deep models in a supervised setting. It also proved that the large quantity of labeled data in supervised approaches could achieve high scores.

Hu and Liu[15][24] introduced the concept of aspect term extraction at implicit and explicit levels. Different models were proposed by the research community with POS and other manually extracted features. Various methods were built considering aspect term extraction as a sequence labeling task and mentioning each word in B-I-O format (Beginning, Inside, or Outside).

Aspect extraction from product reviews has been built using various rules and with the help of SenticNet3 [24]. Authors devised various rules, including subject noun rules, validated on multiple datasets, and reported results. [7][26] applied an unsupervised approach with a contrastive attention mechanism for aspect term extraction. The authors applied POS tagging and in-domain word embeddings trained on a small set of documents. The model utilizes kernel functions to achieve better results in an unsupervised setting.

Tulkens, in [26], devised an unsupervised method to extract aspects using word embeddings and POS tags with the help of spaCy. [28] proposed a neural network architecture POS-AttWD-BLSTM_CRF in combination with minimal features. The authors conclude that the aspect terms are majorly nouns and nouns associated with adjectives. [3] implemented a sequence-to-sequence methodology and I-O-B tag format to extract the aspects from the opinion reviews. The model was adopted from Transformer [27] and applied conditional augmentation for aspect term extraction. [8] used a non-local attention-based convolutional neural model for aspect term extraction. The non-local attention mechanism obtains the dependence features between different words.

2.2 Suggestion Mining

In the fields of NLP and sentiment analysis, suggestion mining is a relatively new task [6]. The majority of the work progressed in suggestion mining, labeling the opinionated sentences as suggestions or not [18][19][20] Viswanathan, Amar, et al. pioneered the notion of suggestion mining into the research aspect for the very first time in the literature. Using rule-based approaches, the writers extracted insights from www.mouthshut.com reviews. The authors tried extracting actionable feedback from the opinion reviews and considered them as suggestions. The authors formulated multiple rules and patterns with linguistic features to extract the suggestions. In [18][19], the authors attempted to use linguistic, n-gram, and POS tag data to identify suggestion expression sentences in customer opinions. The authors also took into consideration both trip advisor and yelp reviews of hotels and devices. The literature lacked a precise definition of suggestion mining until 2015. By compiling information from a variety of restaurants and electronic product reviews, Microsoft windows phone tweets, and software forum conversations, Negi, S. et al., [18][19] available the annotated datasets and problem definition.

In later years, an additional set of reviews was acquired from Twitter and travel-related portals and made available. Using rule-based systems and deep learning methods like LSTM and CNN, the authors attempted to characterize the reviews as suggestions or non-suggestion. The deep models are built by initializing with Word2Vec of 300 dimensions and Glove word embedding of 50- and 100-dimension representations, and it was found that LSTM performed better. The author developed a hybrid system [18][19] to detect the review sentences conveying the suggestion intent. The authors devised a semi-supervised learning method to extract customer-to-customer suggestions from reviews.

To get more research attention and aggressive study on suggestion mining, Sapna Negi et al. organized a pilot task on Suggestion Mining as a part of SemEval-2019 by creating labeled data from feedback forums and hotel reviews. The shared task consists of two subtasks: open-domain and cross-domain suggestion classification. In subtask-1, the model for the classification of opinion reviews is to be trained and evaluated on the same domain of data. In subtask-2, the model training and evaluation must be done on different domain datasets. In response to the pilot task in SemEval-2019, a total of 33 teams participated with their submissions. The majority of submissions are with using pre-trained models and transfer learning approaches.

To describe the results, [9] employed various fundamental machine-learning methodologies. Majorly partitioning techniques such as logistic regression, support vector machines, tree-based models such as decision trees and ensemble approaches, random forest, statistical models, Naive Bayes, and genetic algorithm-based methods are applied. The rule-based system with handmade features is utilized as input to Bi-LSTM models and reported results [2][29]. Most of the additional submissions for these two subtasks centered on pre-trained transfer learning strategies such as BERT and other deep neural network models. In the suggestion mining subtasks, the BERT-based models claimed the top positions.

Dataset	Suggestion		Non-Suggestion	
	Train	Test	Train	Test
Travel Reviews	1050	260	3100	775
MS Windows Phone	1665	420	5135	1280

Jain et al. in [28] built a transformer-based method for the suggestion mining system. To deal with the class imbalance problem, the authors applied Synthetic Minority Oversampling Technique (SMOTE) and Language Model-based Oversampling Technique (LMOTE). The oversampling techniques could boost the performance of the model marginally better. Leekha. M, Goswami, et al. implemented a multi-task learning approach combined with an over-sampling method for suggestion mining. An ensembled RCNN, CNN, and Bi-LSTM with ELMo embeddings are used in multi-task learning. In continuation of SemEval-2019 participation, a few researchers are working on other dimensions of suggestion mining. [10][11] authors built an explainable system with various word embedding and neural networks to identify the learning patterns and understand the significant suggestion components mining.

3. METHODOLOGY

In this section, first, we present the system adopted to deal with the class imbalance problem for suggestion mining. Then, in section 3.2, we elaborate on the reformulation of the problem and its necessity; in section 3.3, the proposed model's overall system architecture and pipeline are explained. Finally, we present the methodology adopted in solving both the sub-problems defined.

3.1 Dataset description

The dataset provided by SemEval-2019 organizers belongs to Travel and Microsoft windows phone reviews. The details of the dataset are presented in Table 1.

Table-1: Statistics of the datasets

3.2 Handling class imbalance

The suggestion mining dataset provided by the SemEval-2019 organizer is imbalanced and biased towards the non-suggestion class. Typically, two approaches are used in the literature to handle class imbalance: data level and algorithm level techniques. The data level approach focuses on up-sampling the minority class to make it near equal to the significant class. In down-sampling, the majority class samples are ignored to meet the minor class number. The down-sampling leads to the loss of labeled data. If sufficient data is unavailable, the deep models do not produce the expected results. Due to this, we employed the up-sampling approach. At the algorithm level, during the neural network training, the computation of loss plays a vital role. In general, cross-entropy has been employed as a loss function for binary classifiers. The focal loss function is the emerging approach in deep learning models to handle the class imbalance nature in data. We utilized a weighted focal loss function to be part of the network training.

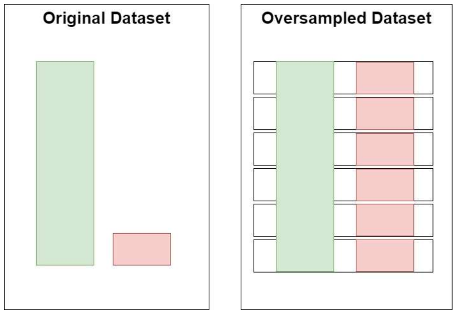

Figure 1: Approach to deal with class imbalance – Oversampling

3.3 Task definition

Given a sentence S ={w_1, w_2, w_3, ……… w_n} with the length n, label the sentence S as a suggestion, non-suggestion. This results in suggestion mining as a binary classification task.

Problem reformulation

The goal of aspect-oriented suggestion mining is not only to classify the given sentence as a suggestion or non-suggestion. It is also required to identify the aspects towards which the suggestions are mentioned in the review sentence. The pioneered research ignored aspect orientation in suggestion mining [18][19]. Fine-grained analysis helps customers to utilize the services better [17][18]. The suggestions help in the quality improvement of the products and services. To achieve this, there is a requirement for labeled data. Due to the non-availability of labeled data for aspect orientation in suggestion mining, we defined it as the two sub-tasks. The sub-tasks are namely Suggestion Aspect Extraction (SAE) and classification. Suggestion aspect extraction aims to identify the aspects in the review sentences. The classification task is to classify the sentences into the various aspects extracted, resulting in a multi-label classification task.

3.3.1 Suggestion aspect extraction

Given a sentence ***S*** = ***{w_1, w_2, w_3, ………w_n}*** of length n, identify the aspects ***{a_1, a_2, a_3, … a_k}*** mentioned in the review, where k is the total aspect categories mentioned in the review.

A review sentence can have multiple aspects towards which the suggestions are mentioned. In preparing the data for model training, each review was mapped with the respective aspect column. In which one indicates review targeting that aspect, zero indicates the absence of suggestion.

For example, in Microsoft Windows reviews, the aspect categories defined are *Features, Tech Support, Windows Specific, Functionality, API, and Miscellaneous*. The aspect terms defined for each category are presented in Table 2.

Table-2: Aspect Categories – Software – Microsoft Windows phone

Aspect Category	Aspect Terms
Features	'language code', 'language', 'calendars', 'validation', 'lock screen', 'synchronize', 'deactivate', 'song collection', 'bluetooth', 'scrollbars', 'file transfer'.
Tech support	'fix', 'bug', 'improve', 'Provide', 'provide', 'introduce', 'implement', 'support', 'incorporate', 'create', 'upload', 'deploy', 'enhance', 'integrate', 'bring','help', 'remove'.
Windows_specific	'microsoft', 'google', 'facebook', 'twitter', 'evernote', 'msdn', 'xbox', 'onedrive'
Functionality	'access', 'messaging', 'image', 'tilting', 'publish', 'build', 'integration', 'brackground', 'pattern'
Miscellaneous	'qr code', 'usb', 'wifi', 'audio', ‘video
API	'api', 'sdk', 'ssl'

The aspects mentioned in Microsoft Windows phone reviews are not suitable for travel reviews. We defined a set of aspect categories and associated aspect terms for travel reviews which are mentioned in Table 3.

Table-3: Aspect Categories – Travel reviews

Aspect Category	Aspect Terms
accessories	'passport', 'passports', 'bag', 'clothes', 'shoes', 'sun glasses',

	'shorts', 'jacket', 'phone', 'camera', 'cash', 'cards', 'laptop', 'adapter', 'charger', 'umbrella', 'travelcard', 'luggage', 'slacks', 'suitcases'
support	'site', 'web', 'register', 'website', 'Research', 'search', 'research', 'picture', 'sticker', 'newspaper', 'tour guide', 'forum', 're-post', 'repost', 'email', 'post', 'ask', 'maps', 'policies', 'video', 'find', 'url', 'try','tag'
places	'lounge', 'restaurant', 'airport', 'hotel', 'beach', 'store', 'food court', 'bank', 'resort', 'shop', 'church', 'ATM', 'bank',
price	'deals', 'fee', 'price', 'rate', 'cost', 'fare', 'money', 'amount', 'refund', 'cheap', '$', 'dollar', 'Euro', 'currency', 'airfare', 'tax', 'discount', 'offer'
Transport	'tour', 'safe', 'cab', 'water', 'taxi', 'tip', 'bus', 'car', 'shuttle','trip'
Miscellaneous	'weather', 'cold', 'tip', 'tips', 'temperature', 'safe',' 'careful','

3.3.2 Classification

Given a sentence $\boldsymbol{S} = \{w_1, w_2, w_3, \ldots\ldots\ldots w_n\}$ of length n, classify the review sentence towards a specific aspect. The class labels $\{C_1, C_2, \ldots\ldots\ldots C_k\}$, are the aspects designated in subtask-1. The goal of subtask-2 is to classify the given sentences into respective aspects.

4 MODEL ARCHITECTURE

A novel system for aspect-oriented suggestion mining from opinion reviews has been proposed in this paper. The overall system of aspect-oriented suggestion mining is presented in the figure below. The workflow consists of two major components: Identifying the aspect terms and defining aspect categories. The second component is the classification model to classify the review sentence towards specific aspects. The models are implemented on suggestion mining labeled data. The suggestion aspect extraction has been applied to identify the aspect terms only on the suggestion class partition of the data. To determine the aspects from review sentences, we adopted a dictionary-based approach to label the aspects for each review. We prepared a dictionary of aspect categories for each domain, such as Software and Travel reviews. For the classification model, along with syntactic features, stylistic and NER features are added to boost the performance of the model.

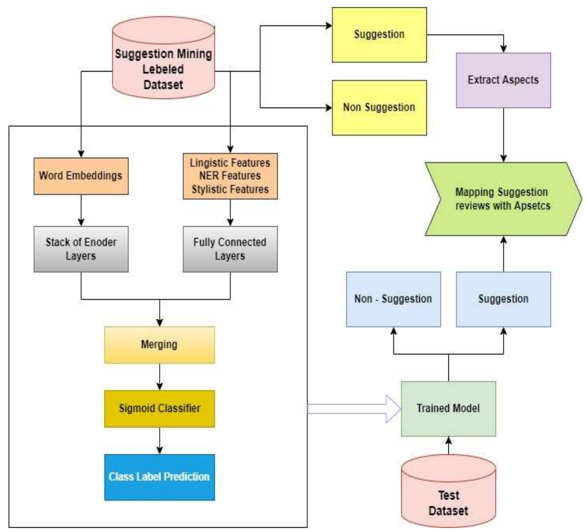

Figure 2: Model Architecture - Aspect-Oriented Suggestion Mining system

5 TASK DESCRIPTION AND EXPERIMENTATION

Suggestion aspect extraction is critical in the entire aspect-oriented suggestion mining pipeline. We implemented an ensemble approach for aspect extraction, with base learning models such as:

- The rule-based approach uses an aspect dictionary
- Entity mapping as an aspect approach
- Computing semantic similarity using cosine similarity and word embeddings

The expert annotated dictionary of aspect categories and aspect terms is used in the rule-based approach. Two such dictionaries are shown in Table 2 and Table 3 for the domains such as Microsoft Windows Phone and Travel reviews, respectively. Step 2 of the aspect extraction algorithm presents the rule-based approach. Each token of the opinion review is compared with the values mentioned in the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

dictionary; if the token is found in values, the respective key represents aspect categories. If more tokens are reporting the steleame aspect category, the same categories are eliminated, and only one time is each category considered in the final set of aspect categories.

In the *Entity mapping approach*, NER has been applied to each review token. The NER tags such as ORG (Organizations), GPE (Geo-Political Entity), PERSON (Names of Persons), and LOC (Location) are considered to claim that the suggestion has been expressed towards the entities as mentioned above, and also the duplicates are eliminated from the list.

In the Semantic similarity-based approach**,** the idea of a dictionary-based approach has been extended to compute the similarity instead of syntactic matching. Each aspect category vector has been represented as the average of the vectors of each aspect word in the dictionary. Each input review is also represented using the vector of the same dimension with a due average of all the word vectors. To identify the aspect for each input review, the cosine similarity is computed among the vector representing the input review and vectors representing aspect categories, and the highest similar category is the final aspect of the review.

In the final step, all the aspects produced by the various approaches mentioned above are merged to give a concise list of all aspects for the single review. The overall process followed for aspect-oriented suggestion mining is depicted below in figure-3 and represented in algorithm-1.

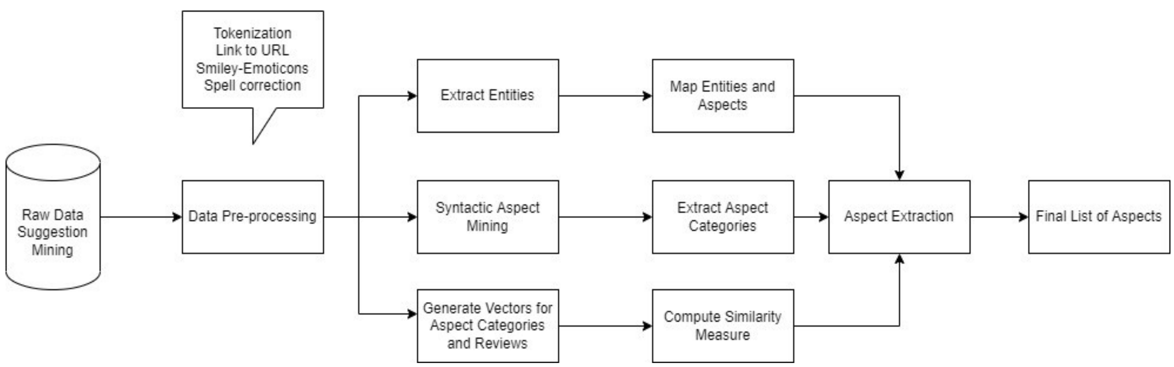

Figure 3: An Ensemble Aspect Extraction Method

5.1.1 Algorithms

Algorithm-1: Procedure for Suggestion Aspect Extraction

Input: Opinion Review statement(s), Dictionary of Aspect Categories and Aspect Terms
Output: Annotated Reviews with Aspects identified
Begin

Step-1
- For each review from the review dataset
  Tokenize each review using spaCy tokenizer
- For each token, identify the entity tag
  If the entity tag is not in the mentioned list
- Consider as an aspect and add to the aspect list
- Return the list of aspects for each review

End step-1

Dataset	Aspects #0	Aspects #1	Aspects #2	Aspects #3	Aspects More than 3
Travel	312	1905	992	329	94
MS Windows Phone	245	1131	658	50	16

Step-2
- For each review from the review dataset
  Tokenize each review using spaCy tokenizer
- For each token, check-in aspect category dictionary, check-in values of the dictionary
- If the value is found, return the key(aspect category) of the value (aspect word)
- Repeat all the tokens in the review sentence and add the aspect categories to the list.

- Return all the aspects

End step-2

Step-3

- Compute the vector representation (R^D) of the aspect category as the average of word vectors (R^D) of the given dictionary
- Compute the vector representation of the review sentence as the average of all word vectors
- Compute the cosine similarity between the review vector and aspect vector
- If the similarity is> 0.4, find the highest similarity value and corresponding aspect category.

End step-3

Step-4:

- Combine the Entity aspects and dictionary-based aspects as the list of all aspects

End step-4

End

The above algorithm generates the aspects for each review. The statistical analysis of the obtained results from the above algorithm is indicated below in Table 4. The Aspect#0 column denotes the number of reviews and has no aspect category words. Similarly, the count of reviews which consists of the number of aspects in each review is mentioned in the below table. The majority of reviews have one or two aspect terms mentioned.

Table-4: Reviews with label count information

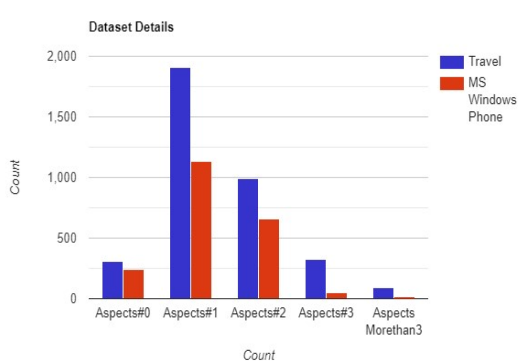

Figure 4: Details of Aspects and Dataset

The findings related to aspect extraction are

- The Named Entity Recognizer (NER) could correctly identify the GPE, LOC, and MONEY aspects.
- The Rule-based Syntactic Aspect Mapping method could correctly identify the aspects of ACCESSORIES, SUPPORT, TRAVEL, and MISC.
- The Word semantic similarity-based approach could correctly identify the ACCESSORIES, SUPPORT, and TRAVEL aspects that are not captured using the syntactic method.
- The combination of the above approach resulted in identifying the aspects correctly.

The devised model may not be efficient in some cases, such as

- When the data has many noise and spelling mistakes
- The approach being rule-based and syntactic may not be efficient for the other domain datasets
- The extensions to the language's vocabulary, the model, may not effectively identify the aspects.

5.2 Suggestion classification

We utilize multiple models from each category of learning algorithms to evaluate the newly articulated aspect-oriented suggestion-mining task. Support Vector Machines (SVM), Naïve Bayes (NB) classifiers as base classifiers, and a one vs. all classifier model built from the machine learning category. Long short-term memory (LSTM) and convolutional neural networks (CNN) are the most popular models for sequence prediction tasks in the neural network category. These two model architectures are used to evaluate the model performance by adding the required number of neurons in the final layer with a softmax classifier. With the advancement of transfer learning and better performance of pre-trained models across NLP tasks, BERT has been employed to validate the results of the task.

5.2.1 Machine Learning approaches

The strategy followed in machine learning approaches for aspect-oriented suggestion mining is fitting one classifier per class. We employed two primary classifiers: SVM and Naïve Bayes. For each classifier, the class is fitted against all the other classes. In addition to its computational efficiency (which required n classifiers only), one advantage of this approach is its interpretability.

- Support Vector Machines is a supervised learning from the computational learning theory, which attempts to find an optimal hyperplane separating two different data classes. An infinite number of hyperplanes can be constructed to separate the data into two classes. However, out of all possible hyperplanes, only one can be constructed with the maximum distance between the hyperplane and the closest data point. That hyperplane is considered to be an optional hyperplane. Such a kind of hyperplane can be identified with a pair of vectors such as weight vector w = w* and bias b = b*. This can be expressed with a combination of two criteria: margin maximization and structural risk minimization. SVMs are popularly used to solve binary classification problems [1][18][19]. However, we consider a multi-label classification problem with a one vs. all approach against each class.
- The Naïve Bayes classifier is the probabilistic classifier, which is based on Baye's theorem. The principle of the Bayes theorem is based on the concept of class conditional independence. The Naïve assumption is that one feature in the data is independent of the other feature present for the same class of data. Majorly two common event models are used for text classification in the literature based on Bayes theorems, such as multinomial naïve bayes and multivariate Bernoulli models. The probabilities are calculated in each of these models by applying Bayes rule to the test data. For example, the probability $P(C_i|x_i)$ is the probability the input X_i belongs to class C_i, likewise, the probabilities for each class are calculated and assigned the label of the class for which the highest probability is given.

5.2.2 Deep Learning Models

Deep Learning models have shown remarkable results across various NLP tasks such as sentiment analysis, text summarization, machine translation, question-answering, image classification, image segmentation, image caption generation, and many more. In this work, we employed two neural network architectures.

- Long Short-Term Memory Network is the popular variant of the recurrent neural network, generally used for sequence prediction tasks. LSTM overcome the limitation of vanilla RNNs, such as exploding and vanishing gradient problems. Thus, LSTM makes it possible to learn the representations in longer sequences [18][19][22]. The internal structure of the LSTM cell consists of three additional neural components called gates to condition the information flow: an input gate, a forget gate and an output gate. These gates help in remembering the previous state's information, forgetting the previous history of hidden states and cells stated, and generating output. This makes it easy way of identifying the aspect class labels for each review by processing through these gates inside each LSTM Cell. The forget gate implemented with the sigmoid function, which outputs the value between zero and one, indicates how much quantity has to be remembered and forgotten. The input layer comprises the tanh function to indicate which values are to be updated. With the help of the gating mechanism in the LSTM network make, remember the long-term dependencies and be capable of sequence prediction.
- Convolutional Neural Networks are popular models for computer vision applications [24][25][30]. Recently, it has been proved that it could also be used for sequence prediction tasks. In this work, we employed CNN to predict the multiple class labels for each review sentence. In CNN, the kernel of the convolution layer slides through the numeric representation of text. The dot-product between the kernel and the embedding of input text is calculated. After that, an activation function is applied to introduce non-linearity into the network. Each kernel produces a feature map when rolled on the input. Then the pooling layer selects the maximum /minimum/average of the feature map values based on the pooling operation type applied to extract the features. The network has

been added with suitable layers at the end of the network based on the kind of output desired.

5.2.3 Transfer Learning Methods

Pre-trained models and transfer learning approaches dominate all tasks across NLP and computer vision. Models like Universal Language Modeling Fine-tuning (ULMFiT), ELMo, BERT, and variants of GPT are used for NLP tasks. Bi-directional Encoder Representation from Transformers (BERT) [8] is the state-of-the-art pre-trained language model architecture based on contextualized word representations and multi-head attention layers. The model is pre-trained with the general-domain corpus, followed by fine-tuning can be done for the specific target task. There are multiple ways to utilize BERT (Devlin et al., 2019), such as suggestion classification. In BERT, different layers capture different levels of semantic and syntactic information. We adopted two models of BERT, such as BERT-base and BERT-large, for multi-label classification to classify the given sentences into aspect classes.

- BERT-base has 12 encoder layers stacked on top of each other. It has 768 hidden layers, along with 12 attention heads and 110 million parameters. We stacked nine dense connections as an additional layer on top of the pooling layer with a sigmoid classifier to classify the output for each aspect.
- BERT large has 24 encoder layers stacked on top of each other. It has 1024 hidden layers along with 16 attention heads and 340 million parameters. Here also stacked nine dense connections as an additional layer on top of the pooling layer with a sigmoid classifier to classify the output for each aspect.

6. RESULT ANALYSIS AND DISCUSSION

Our experimental study was conducted on two datasets: Travel reviews and Microsoft Windows phone reviews. In the travel dataset, there were 5185 reviews, in which 1310 suggestion class, and 3875 non-suggestion classes. Similarly, the Microsoft Windows dataset has 2085 suggestion reviews and 6415 non-suggestions. The ensemble methods for aspect extraction are applied, and the new candidate dataset has been devised, translating the binary classification problem into a multi-label classification task. Table 4 demonstrates the number of aspects considered for the classification of both datasets. Multiple aspects are mapped to each review sentence with the syntactic and semantic approaches. However, the majority of reviews are given with one or two aspects.

Long Short-Term Memory network models are shown remarkable results across various NLP tasks. We considered LSTM as the baseline model for our multi-class classification problem. LSTM models are tested by equipping different word embedding approaches such as random initialization, word2vec, and glove vectors. The results of the F1 score are presented in Table 5. Glove embeddings are capable of sub-word information and better semantic representation of information. Glove outperformed with LSTM network for the aspect-oriented suggestion mining task.

Table 5: Performance analysis Suggestion Class Detection - LSTM

Dataset	Model + Feature	Accuracy
Travel	LSTM + Random	77.7
Travel	LSTM + Word2Vec	78.5
Travel	LSTM + Glove	79.0
MS Windows Phone	LSTM + Random	65.1
MS Windows Phone	LSTM + Word2Vec	66.2
MS Windows Phone	LSTM + Glove	67.4

Journal of Theoretical and Applied Information Technology
30th June 2023. Vol.101. No 12
© 2023 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

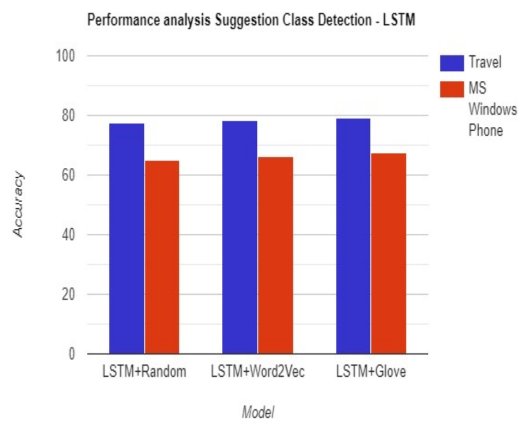

Figure 5: LSTM model performance for Suggestion mining

Table 6 presents the F1 score of the LSTM model with different word embedding combinations. The LSTM with word2vec embeddings could be able to perform better for the aspect-oriented suggestion mining task.

Table-6: Comparison of different embeddings for suggestion mining using LSTM

Dataset	Feature / Embedding	F1-Score
Travel	Random	55.2
Travel	Word2Vec	58.8
Travel	GloVe	58.3
MS Windows Phone	Random	42.5
MS Windows Phone	Word2Vec	51.5
MS Windows Phone	GloVe	64.5

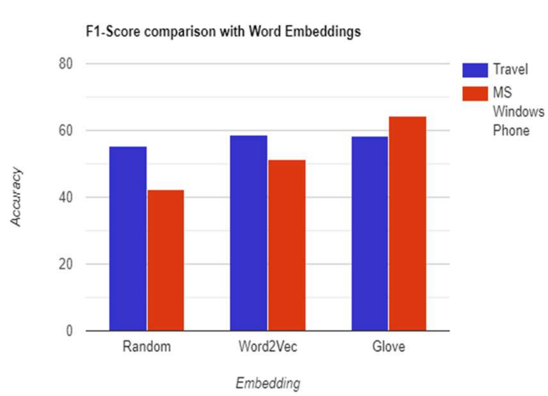

Figure 6: LSTM model performance with Word Embeddings

In combination with the baseline model LSTM, the performance has been evaluated with models such as SVM, Naïve Bayes, CNN, and BERT. Table 7 compares machine learning models' performance with different word embedding combinations. Table 8 demonstrates the comparative study using various word embeddings with deep learning models such as LSTM and CNN. For example, CNN could outperform with a kernel size of 5, which could capture the n-gram feature, enhancing the result of the suggestion mining task. Table- Y demonstrates the performance of pre-trained models such as BERT base and BERT large. We observe that with less availability of labeled data for AOSM and an imbalanced nature of multiple class labels. Pre-trained models could outperform in training, and the model gets overfitting. All the results are presented in multiple tables with different vector initialization methods, and the best results for each pair of methods and evaluation measures are stressed in **boldface**.

Table-7: Comparative analysis of Machine Learning and statistical models for suggestion mining

Dataset	Model	Accuracy	Precision	Recall	F1-Score
Travel	SVM + BoW	41.2	47.1	43.6	45.3
Travel	NB + BoW	57.1	49.1	45.6	45.6
Travel	SVM + TF-IDF	55.3	51.9	52.3	52.5
Travel	NB + TF-IDF	56.3	51.8	50.5	50.7
MS Windows Phone	SVM + BoW	51.8	51.3	44.3	44.4
MS Windows Phone	NB + BoW	49.8	49.5	49.3	49.1
MS Windows Phone	SVM + TF-IDF	51.1	49.4	50.1	48.3

© 2023 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

MS Windows Phone	NB + TF-IDF	50.9	50.9	53.0	50.7

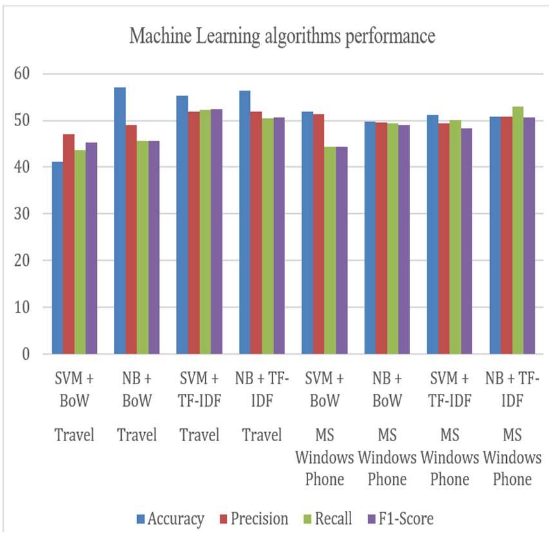

Figure 7: Machine Learning algorithms performance

Table-8: Comparative analysis of Deep Learning Models for suggestion mining on Travel dataset

Model + Features	Accuracy	Precision	Recall	F1-Score
LSTM + Random	51.2	49.1	43.6	45.3
CNN + Random	67.1	48.1	45.6	45.6
LSTM + Word2Vec	55.3	53.9	52.3	52.5
CNN + Word2Vec	66.3	52.8	50.5	50.7
LSTM + Glove	54.2	54.7	54.4	53.9
CNN + Glove	62.2	53.3	51.6	51.3

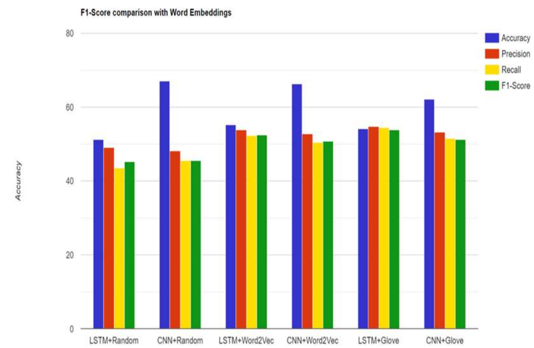

Figure 8: Deep Learning Models performance for Suggestion Mining

The above table and graph demonstrate the performance of two baseline models for deep learning algorithms such as LSTM and CNN for travel datasets. The above models fed with different word embeddings namely random initialization, word2vec, and glove vectors. For the suggestion mining classification task, LSTM with word2vec embedding outperformed.

Table-9: Deep Learning Models performance for suggestion mining on MS Windows Phone dataset

Model + Features	Accuracy	Precision	Recall	F1-Score
LSTM + Random	35.3	28.2	27.7	27.7
CNN + Random	51.8	25.3	24.3	24.4
LSTM + Word2Vec	41.8	29.5	29.3	29.1
CNN + Word2Vec	51.1	26.4	25.1	25.3
LSTM + Glove	40.9	32.9	33.0	32.7
CNN + Glove	47.7	27.1	25.9	26.1

© 2023 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

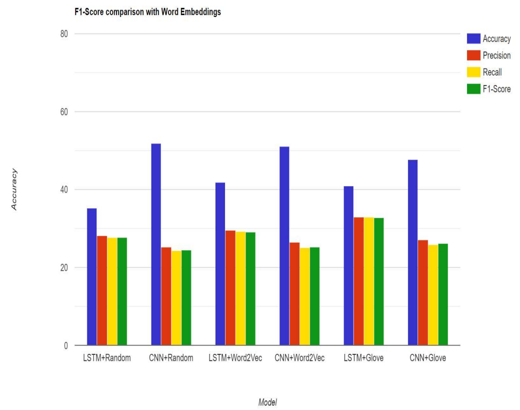

Figure 9 Deep Learning models performance

Table -10: Transfer learning model performance for suggestion mining

Dataset	Model	Accuracy	Precision	Recall	F1-Score
Travel	BERT-base	92.2	89.2	86.3	87.8
Travel	BERT-large	94.5	90.2	89.6	89.6
MS Windows Phone	BERT-base	87.4	82.4	80.6	84.5
MS Windows Phone	BERT-large	88.2	85.3	84.6	85.7

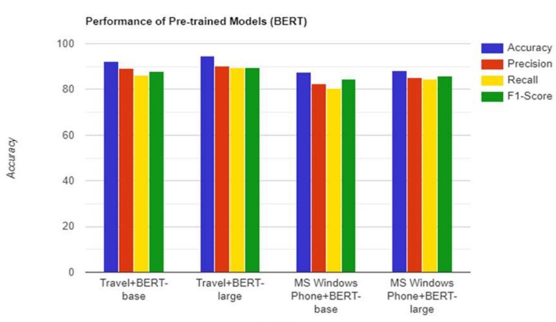

Figure 10 Transfer Learning approaches performance

7. CONCLUSION AND FUTURE SCOPE

This paper presented a hybrid approach toward aspect-oriented suggestion mining from opinion reviews. The approach was devised in two phases: suggestion aspect extraction and suggestion classification. The rule-based and ensemble methods are adopted to aspect term extraction from opinion reviews. The classification of opinion reviews into suggestion or non-suggestion. Along with binary type, classification towards each aspect is also carried with models ranging from Machine Learning to state-of-the-art transfer Learning using Transformer architecture. The transfer learning approach performance is the best among all the models experimented. In the future, we would like to adopt state-of-the-art approaches to extract the exact segment of text, which expresses the suggestion in opinion reviews.

REFERENCES:

[1] *2019 - SemEval-2019 Task 9 Suggestion Mining from Online Reviews and Forums*. (2019).

[2] Baby, P., & B, K. (2020). Sentimental Analysis and Deep Learning : A Survey. *International Journal of Scientific Research in Science, Engineering and Technology*, 212–220. https://doi.org/10.32628/ijsrset207135

[3] Bandorski, D., Kurniawan, N., Baltes, P., Hoeltgen, R., Hecker, M., Stunder, D., & Keuchel, M. (2016). Contraindications for video capsule endoscopy. In *World Journal of Gastroenterology* (Vol. 22, Issue 45). https://doi.org/10.3748/wjg.v22.i45.9898

[4] Bhoi, A., & Joshi, S. (2018). *Various Approaches to Aspect-based Sentiment Analysis*. http://arxiv.org/abs/1805.01984

[5] Bornebusch, F., Cancino, G., Fanseu, A. N., Jalali, M. F., Mohsen, J., Nitze, M., Plump, C., Yotedje, R. S. D., Tchambo, H. F., Toni, & Ziegler, H. (2014). Aspect-based sentiment analysis. *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft Fur Informatik (GI)*, *P-232*, 2389–2400.

[6] Brun, C., & Hagège, C. (2013). Suggestion Mining: Detecting Suggestions for Improvement in Users' Comments. *Research in Computing Science*, *70*(1), 199–209. https://doi.org/10.13053/rcs-70-1-15

[7] Chen, P., Sun, Z., Bing, L., & Yang, W. (2017). Recurrent attention network on memory for aspect sentiment analysis. *EMNLP 2017 - Conference on Empirical Methods in Natural Language Processing, Proceedings*, 452–461. https://doi.org/10.18653/v1/d17-1047

[8] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*(Mlm), 4171–4186.

[9] Fatyanosa, T. N., Hafiz, A., Siagian, A. M., & Aritsugi, M. (n.d.). *DBMS-KU at SemEval-2019 Task 9: Exploring Machine Learning Approaches in Classifying Text as Suggestion or Non-Suggestion*. https://seaborn.pydata.org/generated/seaborn.heatmap.html

[10] Laskari, N. K., & Sanampudi, S. K. (n.d.). *Aspect Based Sentiment Analysis Survey*. *18*(2), 24–28. https://doi.org/10.9790/0661-18212428

[11] Laskari, Naveen Kumar, and Suresh Kumar Sanampudi. "Explainable System for Suggestion Mining using attention." 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS). Vol. 1. IEEE, 2022.

[12] Lin, Y., Wang, X., & Zhou, A. (2016a). Opinion spam detection. In *Opinion Analysis for Online Reviews*. WORLD SCIENTIFIC. https://doi.org/10.1142/9789813100459_0007

[13] Lin, Y., Wang, X., & Zhou, A. (2016b). Opinion spam detection. In *Opinion Analysis for Online Reviews* (Issue May, pp. 79–94). WORLD SCIENTIFIC. https://doi.org/10.1142/9789813100459_0007

[14] Liu, J., Wang, S., & Sun, Y. (n.d.). *OleNet at SemEval-2019 Task 9: BERT based Multi-Perspective Models for Suggestion Mining*.

[15] Liu, P., Joty, S., & Meng, H. (2015). Fine-grained opinion mining with recurrent neural networks and word embeddings. *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing*, *September*, 1433–1443. https://doi.org/10.18653/v1/d15-1168

[16] Manek, A. S., Shenoy, P. D., Mohan, M. C., & Venugopal, K. R. (2017). Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier. *World Wide Web*, *20*(2), 135–154. https://doi.org/10.1007/s11280-015-0381-x

[17] Markov, I., & De La Clergerie, E. V. (n.d.). *INRIA at SemEval-2019 Task 9: Suggestion Mining Using SVM with Handcrafted Features*. https://www.uservoice.com

[18] Negi, S., & Buitelaar, P. (2017). Suggestion Mining From Opinionated Text. In *Sentiment Analysis in Social Networks* (pp. 129–139). Elsevier Inc. https://doi.org/10.1016/B978-0-12-804412-4.00008-5

[19] Negi, Sapna, & Buitelaar, P. (2015). Towards the extraction of customer-to-customer suggestions from reviews. In *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics. https://doi.org/10.18653/v1/d15-1258

[20] Negi, Sapna, & Buitelaar, P. (2017). *Inducing Distant Supervision in Suggestion Mining through Part-of-Speech Embeddings*. http://arxiv.org/abs/1709.07403

[21] Pannala, N. U., Nawarathna, C. P., Jayakody, J. T. K., Rupasinghe, L., & Krishnadeva, K. (2017). Supervised learning based approach to aspect based sentiment analysis. *Proceedings - 2016 16th IEEE International Conference on Computer and Information Technology, CIT 2016, 2016 6th International Symposium on Cloud and Service Computing, IEEE SC2 2016 and 2016 International Symposium on Security and Privacy in Social Netwo*, 662–666. https://doi.org/10.1109/CIT.2016.107

[22] Pecar, S., Simko, M., & Bielikova, M. (2019). *NL-FIIT at SemEval-2019 Task 9: Neural Model Ensemble for Suggestion Mining*. http://arxiv.org/abs/1904.02981

[23] Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*, *108*, 42–49. https://doi.org/10.1016/j.knosys.2016.06.009

[24] Poria, S., Cambria, E., Ku, L. W., Gui, C., & Gelbukh, A. (2014). A Rule-Based Approach to Aspect Extraction from Product Reviews. *SocialNLP 2014 - 2nd Workshop on Natural Language Processing for Social Media, in Conjunction with COLING 2014*, 28–37. https://doi.org/10.3115/v1/w14-5905

[25] Prasanna, S., & Seelan, S. A. (2019). *Zoho at SemEval-2019 Task 9: Semi-supervised Domain Adaptation using Tri-training for Suggestion Mining*. http://arxiv.org/abs/1902.10623

[26] Schouten, K., & Frasincar, F. (2016). Survey on Aspect-Level Sentiment Analysis. *IEEE Transactions on Knowledge and Data Engineering*, *28*(3), 813–830. https://doi.org/10.1109/TKDE.2015.2485209

[27] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

You Need. http://arxiv.org/abs/1706.03762

[28] Yue, P., Wang, J., & Zhang, X. (n.d.). *YNU-HPCC at SemEval-2019 Task 9: Using a BERT and CNN-BiLSTM-GRU Model for Suggestion Mining*.

[29] Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *8*(4), 1–25. https://doi.org/10.1002/widm.1253

[30] Zhou, Q., Zhang, Z., Wu, H., & Wang, L. (2019). *ZQM at SemEval-2019 Task9: A Single Layer CNN Based on Pre-trained Model for Suggestion Mining*. https://github.com/google-research/bert