Aspect-Based Sentiment Analysis Review

Adil Asad
CS Department - Harrisburg Campus
Pennsylvania State University
afa6316@psu.edu

***Abstract*—Aspect based sentiment analysis has become a critical tool for businesses and organizations to understand consumer sentiment and improve their products or services accordingly. In this paper, I review the techniques used for aspect-based sentiment analysis in the last 15 years. My review covers a wide range of approaches, including rule-based methods, machine learning-based methods, and hybrid approaches. I also explore the challenges and limitations of aspect-based sentiment analysis, such as dealing with sarcasm, irony, and other forms of figurative language. Finally, I discuss the current state of the art in aspect-based sentiment analysis and identify promising areas for future research.**

***Index Terms*—aspect-based, sentiment analysis, review**

I. Introduction

The use of public opinion as a decision-making method is well-known in human society, and companies and researchers often use methods like polling and surveys to gather information. With the development of microblogs and e-commerce websites, people are now able to share their opinions on various subjects, and users often refer to these reviews before purchasing a product or using a service. However, the overwhelming number of reviews makes it difficult for users to read and analyze them all. Therefore, there is a need for methods to summarize reviews, and sentiment analysis has become a popular method to do that in recent years.

Sentiment analysis is a rapidly growing field that has garnered significant attention in recent years due to the increasing availability of enormous amounts of user-generated content on social media and other online platforms. One sub-field of sentiment analysis that has received particular attention is aspect-based sentiment analysis, which seeks to identify the sentiment expressed towards specific aspects or features of a product, service, or event. Aspect based sentiment analysis focuses on a particular aspect of the input content. It is a sub-field of sentiment analysis aiming to identify sentiments expressed towards the specific aspect of the product, service, or event's features. Aspect based sentiment analysis has become a critical tool for businesses and organizations to understand consumer sentiment and improve their products or services accordingly.

In this paper, I review the techniques used for aspect-based sentiment analysis in the last 15 years. This review aims to provide a comprehensive overview of the various approaches, methods, and tools developed over the years to perform aspect-based sentiment analysis. In the beginning, I provide a brief introduction to the concept of aspect-based sentiment analysis, its importance, and its applications. I then present a chronological survey of the various techniques proposed and discuss their strengths and weaknesses.

My review covers a wide range of approaches, including rule-based methods, machine learning-based methods, and hybrid approaches. I also explore the challenges and limitations of aspect-based sentiment analysis, such as dealing with sarcasm, irony, and other forms of figurative language. Finally, I discuss the current state of the art in aspect-based sentiment analysis and identify promising areas for future research. Overall, this review provides a comprehensive understanding of the evolution of aspect-based sentiment analysis techniques over the past decades and the current state of the field.

II. Paper Reviews

A. Before 2010

Aspect-based sentence segmentation for sentiment summarization, G. Goldensohn, K. Hannan, 2009.

In the paper, 'aspect-based sentence segmentation for sentiment summarization [2], the authors discuss the importance of aspect summarization when it comes to online reviews, for example when people are considering reviews about a particular product or service, they care about a particular aspect of the item or service. To solve this problem, the authors use an aspect-based sentence segmentation system, which is an unsupervised segmentation model to segment multi-aspect sentences in the data which is a collection of reviews.

In segmentation of sentences, they face two challenges, it is hard to determine topic changes for the purpose of segmentation if you only consider one sentence at a time. Secondly, aspect-based sentence segmentation models need to detect aspect change and polarity changes in the sentence which does not happen in linear text segmentation techniques. To solve these problems, they divide the model into two stages, in the first stage, the input sentence is divided into multiple segments that have only one aspect each and a j score is calculated for each segment to give them an order based on importance. In the second stage of the model, the segments are checked for polarity, if the segments produced in the first stage have more than one polarity, the segments are further divided. For the polarity analysis they use a sentiment-lexicon-based method, in which the semantic orientation value is calculated by adding together the polarity values of all the sentiment words in the segment.

Algorithm: Bootstrapping-based ART Learning
Input: the initial aspect seed sets S={S_1, S_2, ..., S_m} for m aspects, and a pool of unlabeled data U
Stage 1: Candidate ART Extraction
Extract nouns, verbs, adjectives, adverbs and top-n multi-word terms recognized by C-value method from U to form a candidate ART set Ω for bootstrapping.
Stage 2: Bootstrapping Learning
Start learning with the seed set S_i for the i^{th} aspect
Repeat
1. Use Equation (1) to calculate *RlogF* score of each candidate in Ω;
2. Select the candidate with the highest *RlogF* score to augment S_i, and remove it from Ω;

Until the predefined stopping criterion[3] is met.
Output: Final ART sets S for m aspects.

Fig. 1. The bootstrapping learning algorithm.

To evaluate the method, they use a corpus [22] of restaurant reviews which contains 50,000+ sentences and a traditional precision, recall and F1 score metrics are used to evaluate the segmentations. The model performs better than any linear segmentation methods. It proves that for aspect-based segmentation, one sentence cannot provide enough context for the linear segmentation techniques to determine the topic of the sentence. It also shows the effectiveness of using segmentation methods that use polar information.

Methods	P	R	F1
Full-stop-based method	0.68	0.44	0.54
Comma-based method	0.17	0.38	0.24
Dotplotting	0.19	0.39	0.25
C99	0.65	0.44	0.53
Fragkou method	0.45	0.46	0.45
Two-stage segmentation model	**0.69**	**0.56**	**0.62**

Fig. 2. Precision (P), Recall (R) and F1 performance of each method for aspect based sentence segmentation. Each bold number denotes the best performance.

Building a Sentiment Summarizer for Local Service Reviews, S. Goldensohn et al, 2008.

In the paper [1], the problem of making the most use out of the reviews submitted online for products and services is tackled in this paper as well, with the abundance of information available for everything online, it is becoming increasingly challenging to navigate to the useful information.

Methods	WindowDiff
Full-stop-based method	0.21
Comma-based method	0.72
Dotplotting	0.69
C99	0.21
Fragkou method	0.26
Two-stage segmentation model	**0.17**

Fig. 3. WindowDiff values of different methods for aspect-based sentence segmentation. The bold number denotes the best performance.

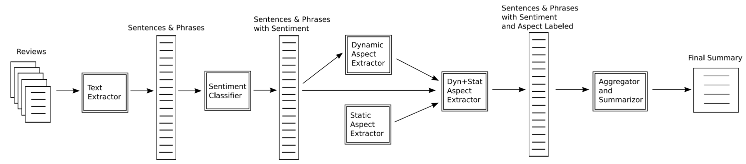

Fig. 4. System Overview. Double boxed items are system components and single boxed items are text files (possibly marked-up with sentiment/aspect information).

The authors of this paper introduce a different approach to summarize opinions on services. A top-level view of their approach can be explained in 3 steps:

1) Identify all sentiment laden text fragments in the reviews

2) Identify relevant aspects for these services that are mentioned in these fragments.

3) Aggregate sentiment over each aspect based on sentiment of mentions

The input for the system in this case is a set of reviews of a particular product or service which is broken into a set of text fragments that can be useful in producing a summary. The fragments are used to aggregate ratings for any aspect mentioned in them. In the second stage all the extracted sentences are classified as either positive, negative, or neutral. For this they employ a hybrid model which uses both machine learning and lexicon-based algorithms. The intent is to show that the context and the global information provided by the user, which can be in the form of a rating, can be used to improve the sentiment classification at sentence level. In the third stage of the system, an aspect is extracted from the input. For this purpose, another hybrid model is used a dynamic aspect extractor, where aspects are determined from the text of the review alone, and a static extractor, where aspects are pre-defined and extraction classifiers are trained on a set of labeled data. Static extractor is a specialized extractor for the domain which helps improve system accuracy.

A problem that is faced when classifying sentences as positive, negative, or neutral is that sometimes-positive reviews can contain some negative opinion and vice versa. So, the sentences are still classified using a model, but that model considers the numerical rating left by the user.

To test the capabilities of the system the authors use a corpus [22] of reviews of different services and an entity to find relevant static and dynamic aspects. The results show that the model produces an f1 score of 75+ for both testing entities. In the last step of the model, a summary is produced about the aspects that the user may be interested in. The authors do not apply a formal evaluation method for the summaries produced by the model.

Multi-aspect opinion polling from textual reviews, J. Zhu et al, 2009.

This paper [3] presents an unsupervised approach to aspect-based opinion polling from textual data. It contains two main components:

	Positive				Negative			
	Precision	Recall	F1	Avg. Prec.	Precision	Recall	F1	Avg. Prec.
raw-score	54.4	74.4	62.9	69.0	61.9	49.0	54.7	70.2
max-ent	62.3	76.3	68.6	80.3	61.9	76.7	68.5	71.3
review-label	63.9	89.6	74.6	66.2	77.0	86.1	81.3	76.6
max-ent-review-label	**68.0**	**90.7**	**77.7**	**83.1**	**77.2**	**86.3**	**81.4**	**84.4**

Fig. 5. Sentiment Classification Precision, Recall, F1, and Average Precision. Systems above the line do not use any user provided information. Bolded numbers represent the best result.

Algorithm 1: Multi-Aspect Bootstrapping Learning
Input: initial aspect seed sets $S=\{S_1, S_2, \ldots, S_m\}$ for *m* aspects, and pool of unlabeled data *U*
Stage 1: Candidate ART Extraction
Extract nouns, verbs, adjectives, adverbs and top-*n* multi-word terms recognized by C-value method from *U* to form a candidate ART set Ω for bootstrapping.
Stage 2: **Single-Aspect Bootstrapping Learning**
Start learning with the seed set S_i for the i^{th} aspect;
Repeat
1. Use Equation (1) to calculate *RlogF* score of each candidate in Ω;
2. Select the candidate with the highest *RlogF* score to augment S_i, and remove it from Ω;

Until the predefined stopping criterion[2] is met.
Stage 3: **Re-scoring**
For each ART set S_i produced by SAB
1. Use Equation (4) to calculate *MAB* score of every ART in S_i;
2. Sort ARTs in the descending order of *MAB* score to produce final S_i.

Output: Final generated ART sets S^* for *m* aspects.

Fig. 6. The multi-aspect bootstrapping algorithm

A multi aspect bootstrapping algorithm to learn aspect related terms from unlabeled data

An unsupervised segmentation model to address the challenge of identifying multiple single aspect units in a multi aspect sentence

To extract multi aspect terms from the data C value method, which is calculated using the formula:

Cvalue(t) = $log(|t|) * frq(t)$

Where $|t|$ denotes the number of words contained by t.

The 'bootstrapping' method used in this paper starts learning with a small number of aspect related terms from the unlabeled data but then performs clustering iteratively, where in each cycle the most valuable candidate from the cycle is chosen to add to the current seed and it continues until the stopping criteria is met.

The sentence segmentation in this project is done using an unsupervised approach, as most review sentences contain multiple aspects, sentence segmentation is important to implement. The model (MAS) takes sentences with multiple aspects and outputs segments with single aspects. Each output segment is scored using the formula:

U* = argmaxJ(C,U)

The aspect score is calculated by summing the importance scores of all the aspect terms and then arranged in order based on score.

An aspect-based opinion poll is described as a two-dimensional form filled with 3m (aspect, polarity) pairs and

Methods	Precision (%)
MAB+MAS	75.5
MAB+Full-Stop	70.3
SAB+MAS	71.2
SAB+Full-Stop	66.4

Fig. 7. Figure shows the effectiveness of each automatic method for aspect-based opinion poll generation. MAB+MAS method achieves the highest average precision of 75.5use labeled data. Experimental results show that in this task MAB outperforms SAB, and MAS is better than full-stop-based method.

their corresponding numbers, involving m aspects and three polarities (i.e., positive, negative, and neutral). For each (aspect, polarity) pair, its associated number indicates how many textual reviews express it.

The method is evaluated on a corpus of reviews available online, for the c value method, bigram, trigram, 4-gram were considered which produced an accuracy of around 75 percent. This proves the effectiveness of automatic methods for aspect-based opinion poll generation.

B. 2010-2014

Lexicon-Based Methods for Sentiment Analysis, M. Taboada et al, 2011.

In the paper [4] a lexicon-based approach is applied to the problem of extracting sentiments from the text. A technique called semantic orientation is used to find the subjectivity measure in the text. Semantic orientation is used to calculate the factor which can be positive or negative and the strength of the word, phrase, segment, or sentence towards the target which can be a topic subject, a person, or an idea. The authors choose to follow a lexicon-based approach to extract the sentiment from the text over a supervised classification approach. They prefer this method of over the classifiers because the classifiers are domain specific, even though they perform well for the domain they are trained for, their performance drops significantly when used in other domains.

To create the dictionary for the semantic orientation calculator, it is necessary to assume that every word has a prior polarity independent of any context and semantic orientation can be expressed as a numerical value. To build the system for their project the authors choose to use the corpus[1] which contains 400 text collection of reviews that are extracted from eight distinct categories of products. The reviews are then split into 25 positive and negative for each category. In total the corpus contains around 280,000 words. To evaluate the performance of their dictionary the authors used the original collection of 400 reviews, another collection of 400 texts from the website epinions.com, a 1900 movie review text dataset[2] and 2400 text corpus of reviews taken from a large set of reviews[3]. The results of the evaluation were obtained by comparing the output of the semantic orientation calculator with the 'recommended' or 'not recommended' section of the review.

Next, the authors compare the performance of their semantic orientation dictionary with the other dictionaries available. The

Dictionary	Percent correct by corpus				
	Epinions 1	Epinions 2	Movie	Camera	Overall
Google-Full	62.00	58.50	66.31	61.25	62.98
Google-Basic	53.25	53.50	67.42	51.40	59.25
Maryland-Full-NoW	58.00	63.75	67.42	59.46	62.65
Maryland-Basic	56.50	56.00	62.26	53.79	58.16
GI-Full	68.00	70.50	64.21	72.33	68.02
GI-Basic	62.50	59.00	65.68	63.87	64.23
SentiWordNet-Full	66.50	66.50	61.89	67.00	65.02
SentiWordNet-Basic	59.25	62.50	62.89	59.92	61.47
Subjectivity-Full	72.75	71.75	65.42	77.21	72.04
Subjectivity-Basic	64.75	63.50	68.63	64.83	66.51
SO-CAL-Full	80.25	80.00	76.37	80.16	**78.74**
SO-CAL-Basic	65.50	65.25	68.05	64.70	66.04

Fig. 8. Comparison of performance using different dictionaries with SO-CAL.

evaluation is done based on the accuracy of the prediction of the polarity of the words by comparing it with the user selected 'Recommended' or 'Not recommended' option of the review.

The results show that a manually built dictionary provides a good base for a lexicon-based approach. Lexicon based models for sentiment analysis are not only robust as they can be applied to multiple domains, but they are also easy to enhance by adding additional sources of knowledge without the need for further development.

Aspect Based Sentiment Analysis using Support Vector Machine Classifier, R. Varghese et al, 2013.

This paper aims to implement aspect-based sentiment analysis using support vector machines classifier. The proposed technique has multiple steps, the first one is classifying the reviews as useful or not. For this implementation, the authors consider a 'useful' review as one that contains an opinion about the product, if there is no opinion shared in the review it is not useful for the model. For example, if there is a question asked by the reviewer during the review process, it contains no opinion, such sentences are found in the reviews and removed. This is done using 'Sentiwordnet' which has a collection of opinion words. If the review does not contain any opinion words or any of the extracted features of the product, it is not considered. For the aspect expression identification, the authors use a technique in which they simultaneously search for proper nouns and phrases or words that have a feature of a product. This is done by part of speech tagging, used to identify nouns and phrases. In most cases the noun phrases are identified to be the features of the product. The authors further use conference resolution to resolve all the anaphora that are there in the review sentences. For this they use 'Stanford Deterministic Coreference Resolution System.' The next step of their process is to extract opinion words for aspects. For that they use Standford dependency parser, which extracts all the direct and transitive dependencies in the text. The next step of their technique is to use a support vector machine classifier. The SVMs are used to create hyperplanes that separate the cases which belong to different classes. The SVM classifier assigns a polarity value to the extracted opinion of the product. The authors then use the result of the SVMs and store it in the form of tuples,

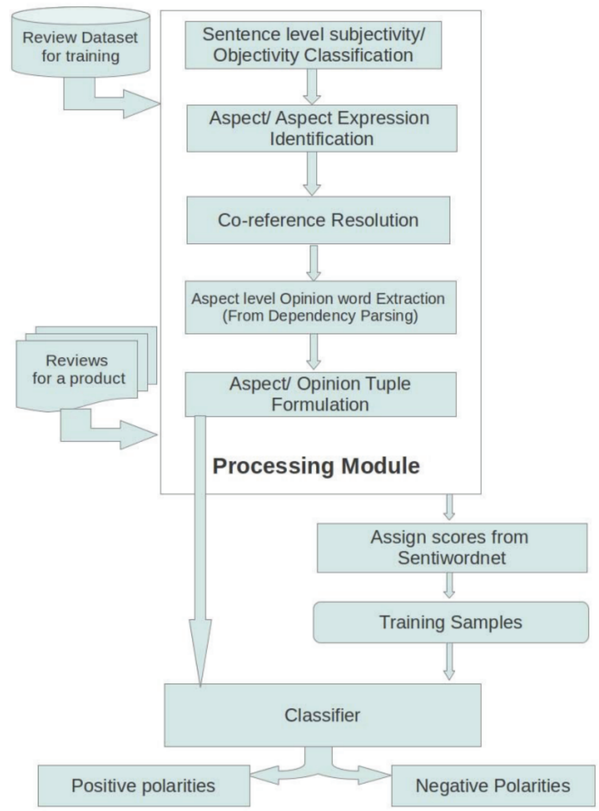

Fig. 9. Structure of the model used by R. Varghese et. al.

P_orient = (r_i,a_{ij},pol_{ij})

Where r_i is the ith review, a_{ij} is the feature j in review I, pol_{ij} is the polarity.

After identifying the polarities of the opinions in all the reviews, they are aggregated using the formula:

$orient\_aggregate$[j] = $\Sigma_i pol_{ij}$

And then normalized using the formula:

$orient\_norm$[j] = $orient\_aggregate$[j] / Σ_i

To evaluate their works results, the authors use the precision, recall and accuracy rates on sentence basis. They test it on a dataset of reviews of digital cameras which are already annotated as positive and negative. They achieve an accuracy of 77.98 percent in their work.

Aspect-Based Twitter Sentiment Classification, H. Lek et al, 2013.

In this paper H. Lek et al. apply aspect-based sentiment analysis to a twitter dataset. Previously there have been multiple applications of sentiment analysis on twitter data but most of these applications consider tweet level sentiment. Applying aspect-based sentiment analysis to tweet data solves

Product	Precision	Recall	Accuracy
Camera 1	85.94%	87.30%	78.48%
Camera 2	93.30%	77.38%	74.82%
Camera 3	91.30%	84.00%	80.65%
Average	90.18	82.89	77.98

Fig. 10. Evaluation Results

a range of problems as a single tweet can have multiple aspects. Getting a tweet level positive or negative sentiment is not very useful in most cases for example a company would be more interested in a particular aspect of the product that the customer finds useful or dislikes than just the fact that they like it or not. This information can be used to make changes and improve the product.

There are three stages of the model the authors applied:
1) Aspect sentiment extraction
2) Aspect ranking and selection
3) Aspect classification

The aspect-sentiment extraction process involves identifying aspects and their associated sentiments and polarity in a tweet. Each noun in a tweet is considered an aspect candidate. A part of speech tagger, sentiment lexicon, and gazetteer lists are used to obtain the results in the form of [aspect, sentiment words, polarity]. The extractor employs the Twitter part of speech tagger, an opinion lexicon, and stop, swear, and intensifier word lists. The extraction process involves tokenization, part of speech tagging, and sentence extraction, considering only the tokens after the coordinating conjunction "but" finding the closest verb and adjective/adverb/hashtag to the left and right of the aspect candidate, considering negation cases, and correcting the word progressively to find a match.

In the aspect ranking and selection step a query is used to retrieve tweets about a particular topic or interest. Then the algorithm from the previous step is applied to the collected tweet data, and the results combined. The number of occurrences of each aspect is summed and then arranged in a descending order. A threshold value is used to see if any aspect falls below it, if it does, it is ignored. Then an approach called Pointwise Mutual Information (PMI) is applied which uses the formula:

$PMI(p, q) = log_2(hits(qANDq)/hits(p).hits(q)$

Where hist(query) is the number of hits returned by the google search for a given query.

Next aspect classification step is applied to find aspects found in the set of important aspects identified in the previous step. And the number of times that particular aspect is classified as positive or negative is counted and an overall label is assigned, if number of positive classifications ¿ number of negative classifications, it is assigned a positive label and vice versa. In the case where the positive and negative classifications are equal, that aspect is ignored.

As the approach applied by H. Lek et al. Is very dependent on lexicons, they propose a technique to improve it. They propose target dependent sentiment expressions extraction. For example,

Sentiment Expressions	Polarity
thank-TARGET	Positive
TARGET-ftw	Positive
TARGET-COP-amazing	Positive
hang-out-with-TARGET	Positive
sad-TARGET	Negative
will-miss-TARGET	Negative
TARGET-screw-up	Negative
so-sorry-to-hear-about-TARGET	Negative

Fig. 11. Examples of sentiment expresions

The expressions in the list are a combination of target and verb phrases. It is performed by using a large corpus that has labels and polarity predefined. The expressions extracted are then passed through a polarity assessment which uses the formulas:

$C_t(e) = C_p(e) + C_n(e)$
$P(positive|e) = C_p(e)/C_t(e)$
$polarity(e) = C_t(e) >= M1andP(positive|e) >= 0.7$; polarity positive
$C_t(e) >= M1andP(positive|e) <= 0.3$; polarity negative
x otherwise

Where Cp(e) is the number of times the expression e appears in the positive tweets and Cn(e) is the number of times the expression appears in the negative tweets.

Since tweets can have multiple expressions that are extracted, these expressions can have similar polarity and can occur in the tweets together, and the opposite polarity of these expressions can also be determined using negation. A graph can be constructed with the expressions as nodes and co-occurrences and polarity as edges. Affinity propagation clustering is used to group nodes into clusters of similar polarity. A simple probabilistic approach is found to be more accurate than the graph-based method for polarity assignment. Out of 48848 sentiment expressions, 12805 are classified as positive and 11642 as negative. The distribution of N-grams of the expressions is shown in table

N-Gram	1	2	3	4	>=5
All	8624	22291	13561	3635	737
Pos/Neg	3747	10592	7452	2157	499

Fig. 12. Distribution of N-grams of the expressions

To evaluate the approach four datasets were used, Stanford Twitter Sentiment (STS), Sanders Twitter Corpus (STC), Telecommunication Company Dataset 1 and 2 (TCD1, TCD2) the tables below show the accuracy of the sentiment analysis techniques proposed on each of the datasets.

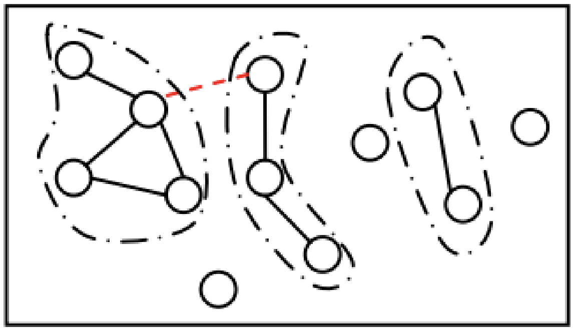

Fig. 13. An example graph of the sentiment expressions

Method	STS	STC
Go et al. [8]	83	-
Speriosu et al. [21]	84.7	-
Saif et al. [20]	86.3	-
MaxEnt	84.7	77.7
MaxEnt+Layered	87.7	**80.2**
MaxEnt+Layered+Ngram	**88.3**	**80.2**
NB	84.1	75.1
NB+Layered	86.3	79.1
NB+Layered+Ngram	87.2	79.3

Fig. 14. Sentiment classification accuracy on STS

Method		P	R	F
Known Aspect	Aspect	**83.6**	75.4	79.3
	Aspect+Ngram	83.3	76.9	**80.0**
	Lexicon	73.2	**80.8**	76.8
	Lek and Poo [12]	75.2	25.1	37.6
Unknown Aspect	Aspect	68.3	70.4	69.3
	Aspect +Ngram	68.3	71.5	**69.8**
	Lexicon	57.9	**76.9**	66.0
	Lek and Poo [12]	**70.6**	20.9	32.3

Fig. 15. Sentiment classification accuracy on TCD1

Method		P	R	F
Known Aspect	Aspect	**88.4**	76.5	82.0
	Aspect+Ngram	88.2	77.9	**82.7**
	Lexicon	81.5	**82.7**	82.1
	Lek and Poo [12]	69.1	26.5	38.3
Unknown Aspect	Aspect	68.8	67.8	68.3
	Aspect+Ngram	**68.9**	68.9	**68.9**
	Lexicon	54.7	**75.4**	63.4
	Lek and Poo [12]	56.2	22.1	31.8

Fig. 16. Sentiment classification accuracy on TCD2

The experimental results show that a layered classification approach is more effective than a classifier trained using target-dependent features.

C. 2015-2019

Aspect Based Sentiment Analysis with Gated Convolutional Networks, W. Xue et al, 2018. In the paper 'Aspect based sentiment analysis with gated convolutional networks' W. Xue et al. propose a model that uses convolutional neural networks with gated mechanisms which increases the accuracy with which it can do aspect based sentiment analysis as compared to previous approaches.

The model they propose uses convolutional layers and gating units. Where each filter computes n-gram features at different granularities from the embedding vectors at each position. The gating units present in the layers at various positions are also independent, therefore the model is more suitable for parallel computing which decreases the time for training. They use a CNN (Convolutional Neural Networks) model which consists of an embedding layer, a convolutional layer, and a max pooling layer. The embedding layer is initialized with pretrained embeddings. The convolutional layer takes the input and outputs multiple convolutional kernels of different widths. Each kernel is a feature detector which is used to extract specific patterns of n-grams.

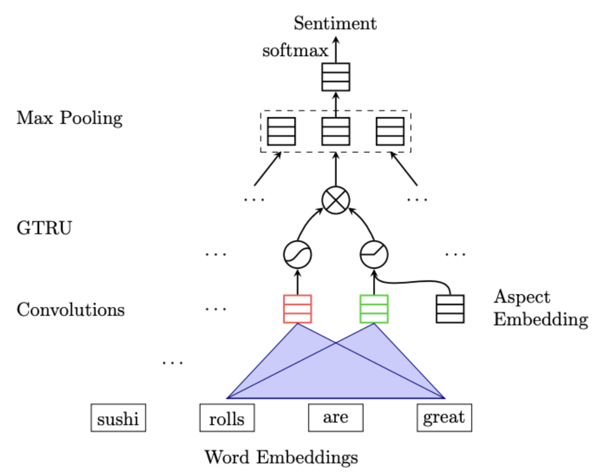

Fig. 17. Illustration of the model GCAE for ACSA task.

The ReLU gate in the figure receives the aspect information to control the propagation of sentiment features. The output from the gates is elementwise multiplied for the max pooling layer. This is fed into two convolutional neurons where features are computed using the following formulas:

$$a_i = relu(X_{i:i+k} * W_a + V_a v_a + b_a)$$
$$s_i = tanh(X_{i:i+k} * W_s + b_s)$$
$$c_i = s_i * a_i$$

Where Va is the embedding vector of the given aspect category. This produces a fixed sized vector e which contains the most prominent features of the sentence. The final layer of the model that has softmax function is used to predict the sentiment polarity. The model is trained by reducing the entropy loss between the predicted and the actual value.

$$L = \Sigma_i \Sigma_j y_i log(yi)$$

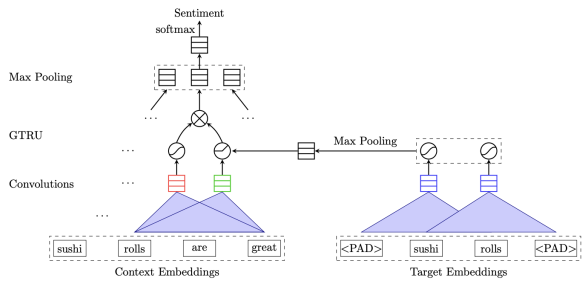

Fig. 18. Illustration of model GCAE for ATSA task. It has an additional convolutional layer on aspect terms.

To test the proposed model for aspect-based sentiment analysis, they used the publicly available datasets from the SemEval workshop that contains reviews for restaurants and laptops. They also created another dataset that contains sentences with opposite sentiments on distinct aspects as datasets containing reviews usually contain different opinions on different aspects of the product so in order to test the model properly this dataset is important. The preexisting models that they use to compare the performance of their model are, NRC-Canada[15], CNN[16], TD-LSTM[17], ATAE-LSTM[18], IAN[19], RAM[20] and GCN.

The results are as follows:

Models	Restaurant-Large		Restaurant 2014	
	Test	Hard Test	Test	Hard Test
SVM*	-	-	75.32	-
SVM + lexicons*	-	-	**82.93**	-
ATAE-LSTM	83.91±0.49	66.32±2.28	78.29±0.68	45.62±0.90
CNN	84.28±0.15	50.43±0.38	79.47±0.32	44.94±0.01
GCN	84.48±0.06	50.08±0.31	**79.67±0.35**	44.49±1.52
GCAE	**85.92±0.27**	**70.75±1.19**	79.35±0.34	**50.55±1.83**

Fig. 19. Results of different models and the proposed model.

Aspect-Based Sentiment Analysis Using BERT, M. Hoang et al, 2018.

This paper 'Aspect-Based Sentiment Analysis Using BERT' proposes a new aspect-based sentiment analysis task for out-of-domain classification at both sentence and text levels and proposes a general classifier model that uses BERT as the base for contextual word representations. The model uses the sentence pair classification model to find semantic similarities between a text and an aspect and outperforms previous submissions, except for one. The paper also proposes a combined model that uses one sentence pair classifier model from BERT to solve both aspect and sentiment classification simultaneously.

The authors first introduce the pre-trained language model used in the paper, Bidirectional Encoder Representations from Transformers (BERT), which has achieved state-of-the-art results in several Natural Language Processing (NLP) tasks. BERT considers the context of a word from both the left and right side simultaneously, which improves results at several NLP tasks. They then explain the Aspect-Based Sentiment Analysis (ABSA) task from SemEval-2016 and describe previous works with and without a pre-trained model.

The authors describe how BERT works by using a masked language model (MLM) to mask a random word in a sentence with a small probability, replace it with a token [MASK], and later try to predict the masked word by using the context from both the left and right of the masked word with the help of transformers. BERT also has an additional key objective, namely next-sentence prediction. The text input for the BERT model is processed through a method called 'wordpiece' tokenization and is later processed through three different embedding layers with the same dimensions that are later summed together and passed to the encoder layer: Token Embedding Layer, Segment Embedding Layer, and Position Embedding Layer.

They also explain that the architecture of transformers is not based on Recurrent Neural Networks (RNNs) but on attention mechanics, which decides what sequences are important in each computational step. The chapter then introduces the Sentence Pair Classifier Task, which deals with determining the semantic relations between two sentences and evaluates how good a model is on comprehensive understanding of natural language and the ability to do further inference on full sentences. Finally, the chapter describes the benchmark that evaluates natural language understanding on models named general language understanding evaluation (GLUE), which consists of several tasks such as multi-genre natural language inference (MNLI), the semantic textual similarity benchmark (STS-B), and Microsoft research paraphrase corpus (MRPC). The paper presents three models for aspect-based sentiment analysis: an aspect classification model, a sentiment polarity classification model, and a combined model for both aspect and sentiment classification. The dataset used is from SemEval-2016, which is pre-processed to fit BERT. The models are trained to predict sentiment labels (positive, negative, neutral, conflict) for a given aspect and text input, and to determine whether the aspect is related to the text or not. The combined model outputs a sentiment label if the aspect is related and the unrelated label otherwise. The unbalanced dataset is weighted to compensate for the imbalance.

	Sentence Level			Text Level		
	Restaurant	Laptop	Both	Restaurant	Laptop	Both
Texts	2000	2500	4500	334	395	729
Unique Aspects	12	81	93	12	81	93
Aspects with Sentiment	2507	2908	5415	1435	2082	3517
Aspects without Sentiment	21493	199592	413085	2573	29913	64280
Total Aspects	24000	202500	418500	4008	31995	67797

Fig. 20. Distribution of data in each training dataset.

The article presents an evaluation of models for aspect categorization and sentiment polarity classification in different domains and levels of text (sentence-level and text-level). The evaluation is based on SemEval-2016 Task 5, and the results are presented in tables, with the best performing models for each type, domain, and level. The aspect classifiers perform better on text-level datasets, while the combined classifiers perform better on sentiment classification tasks. Out-of-scope evaluations show that classifiers trained on datasets with more unique aspects or sentence-level datasets perform better.

Model	Domain	Level	F1	PRE	REC	ACC
ASP	REST	SENL	**79.9**	80.2	79.5	96.3
COM	REST	SENL	77.4	75.9	79.0	95.8
ASP	REST	TEXL	55.5	41.0	85.9	87.4
COM	LAPT	SENL	35.7	30.0	44.1	85.5
Baseline: BERT-PT			78.0	-	-	-
Baseline: NLANGP			73.0	-	-	-

(a) Results of Aspect models on dataset: Restaurant, Sentence-Level. BERT-PT (Xu et al., 2019) and NLANGP (Toh and Su, 2016) as baselines

Model	Domain	Level	F1	PRE	REC	ACC
ASP	BOTH	SENL	**51.7**	40.7	70.6	98.4
ASP	BOTH	TEXL	39.0	27.5	66.7	97.5
COM	BOTH	SENL	38.7	25.5	80.7	96.9
ASP	REST	SENL	5.7	3.0	67.3	73.5
Baseline: NLANGP			**51.9**	-	-	-

(b) Results of Aspect models on dataset: Laptop, Sentence Level. With NLANGP (Toh and Su, 2016) as baseline.

Model	Domain	Level	F1	PRE	REC	ACC
ASP	BOTH	SENL	**34.4**	23.3	65.9	89.1
COM	REST	SENL	34.1	22.9	67.5	88.7
ASP	LAPT	TEXL	33.8	28.2	42.1	92.8

(c) Performance of Aspect models on the dataset: Hotel, Sentence-Level.

Fig. 21. Best performance of aspect category classifiers in sentence-level datasets

Model	Domain	Level	F1	PRE	REC	ACC
ASP	REST	TEXL	**85.0**	84.2	85.9	88.7
COM	BOTH	TEXL	82.4	78.2	87.1	86.1
ASP	BOTH	SENL	78.8	81.9	76.0	84.7
COM	LAPT	TEXL	68.0	66.4	69.6	75.5
Baseline: GTI			84.0	-	-	-

(a) Results of Aspect models on dataset: Restaurant, Text-Level. Baseline: GTI (Álvarez-López et al., 2016).

Model	Domain	Level	F1	PRE	REC	ACC
ASP	BOTH	TEXL	**64.3**	60.9	68.1	92.3
COM	BOTH	TEXL	63.9	57.4	72.1	91.7
ASP	LAPT	SENL	61.0	58.7	64.6	91.6
COM	REST	SENL	21.6	12.3	87.4	37.0
Baseline: UWB			60.5	-	-	-

(b) Performance of Aspect models on the dataset: Laptop, Text-Level. UWB (Hercig et al., 2016) as baseline.

Model	Domain	Level	F1	PRE	REC	ACC
ASP	BOTH	TEXL	**60.8**	48.3	82.0	62.8
COM	LAPT	TEXL	59.4	53.7	66.4	68.0
COM	BOTH	SENL	58.8	45.2	84.0	58.5
ASP	REST	SENL	56.7	45.0	76.7	58.8

(c) Results of aspect category classifiers on the dataset: Hotel, Text-Level

Fig. 22. Best performance of aspect category classifiers in text-level datasets

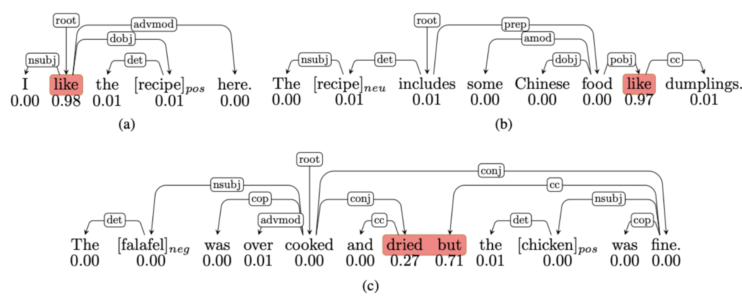

Fig. 23. Three examples from restaurant reviews to illustrate the relationships among aspect, attention, and syntax in ABSA. Labeled edges indicate dependency relations, and scores under each word represent attention weights assigned by the attention-equipped LSTM. Words with high attention weights are highlighted in red boxes, and words in brackets are the target aspects followed by their sentiment labels.

D. 2020-2023

Relational Graph Attention Network for Aspect-based Sentiment Analysis, K. Wang et al, 2020.

In the paper 'Relational Graph Attention Network for Aspect-based Sentiment Analysis' the authors discuss recent efforts have used attention mechanisms to implement aspect-based sentiment analysis, but they occasionally fail. Other approaches explicitly leverage the syntactic structure of a sentence, but they have limitations, such as ignoring dependency relations and encoding unnecessary parts of the parse tree. To address these issues, the authors propose a novel aspect-oriented dependency tree structure and a relational graph attention network (RGAT) model to encode the dependency relations and establish the connections between aspects and opinion words. The proposed approach achieves superior performance to baseline methods on the SemEval 2014 and Twitter datasets.

The authors discuss the limitations of attention-based models in aspect-based sentiment analysis (ABSA) and propose a novel aspect-oriented dependency tree structure to address these limitations. The proposed structure reshapes an original dependency tree to root it at a target aspect, followed by pruning of the tree to discard unnecessary relations. The structure has two advantages: each aspect has its own dependency tree and can be less influenced by unrelated nodes and relations, and if an aspect contains more than one word, the dependency relations will be aggregated at the aspect. The proposed approach provides a direct way to model the context information and enables batch and parallel operations during training.

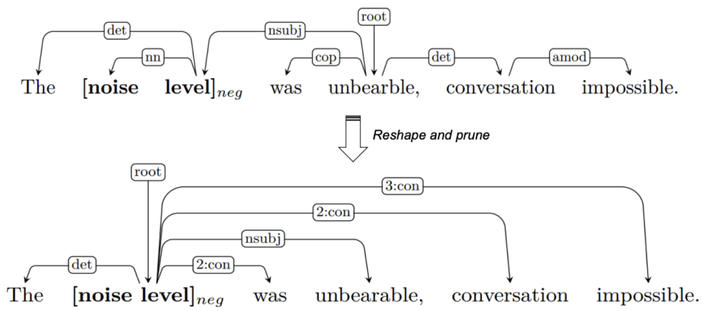

Fig. 24. Construction of an aspect-oriented dependency tree (bottom) from an ordinary dependency tree (top).

Algorithm 1 Aspect-Oriented Dependency Tree

Input: aspect $a = \{w_i^a, w_{i+1}^a, ...w_k^a\}$, sentence $s = \{w_1^s, w_2^s, ...w_n^s\}$, dependency tree T, and dependency relations r.

Output: aspect-oriented dependency tree $\hat{T}$.

Construct the root R for $\hat{T}$;
for i to k **do**
  for $j = 1$ to n **do**
    if $w_j^s \xrightarrow{r_{ji}} w_i^a$ **then**
      $w_j^s \xrightarrow{r_{ji}} R$
    else if $w_j^s \xleftarrow{r_{ij}} w_i^a$ **then**
      $w_j^s \xleftarrow{r_{ij}} R$
    else
      $n = distance(i, j)$
      $w_j^s \xrightarrow{n:con} R$
    end if
  end for
end for
return $\hat{T}$

Fig. 25. Algorithm for an aspect oriented dependency tree

The paper proposes a Relational Graph Attention Network to encode dependency trees for sentiment analysis. The Relational Graph Attention Network is an extension of the Graph Attention Network that encodes graphs with labeled edges. Graph Attention Network iteratively updates each node's representation by aggregating neighborhood node representations using multi-head attention. However, this process fails to account for dependency relations, leading to the loss of valuable information. The Relational Graph Attention Network extends the Graph Attention Network with additional relational heads that act as relation-wise gates to control information flow from neighborhood nodes. The final representation of each node is computed by concatenating attentional heads and relational heads. The authors use BiLSTM to encode word embeddings of tree nodes and obtain the output hidden state for the initial representation of the leaf node. After applying Relational Graph Attention Network to an aspect-oriented tree, its root representation is passed through a fully connected softmax layer and mapped to probabilities over different sentiment polarities. The authors use the standard cross-entropy loss as their objective function.

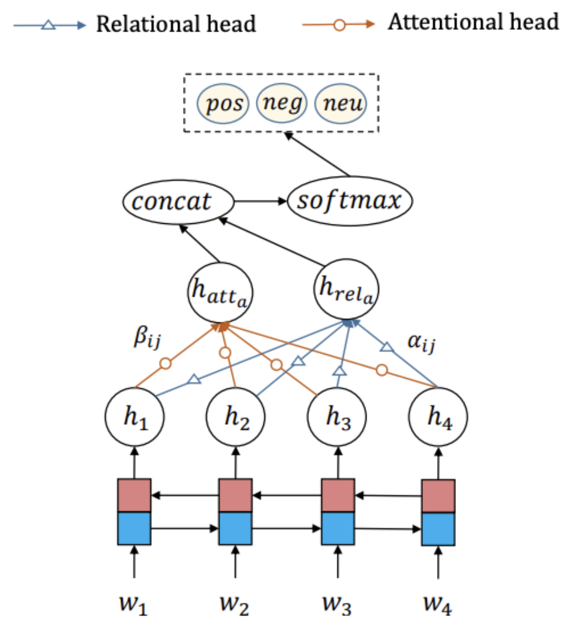

Fig. 26. Structure of the proposed relational graph attention network (R-GAT), which includes two genres of multi-head attention mechanism, i.e., attentional head and relational head.

Three public datasets (Laptop, Restaurant and Twitter) are used for evaluation and comparison with existing mainstream models. The model is shown to outperform most of the baseline models, with the basic BERT model already outperforming all the existing aspect-based sentiment analysis models by significant margins. The combination of R-GAT and BERT achieves a new state of the art. These results demonstrate the effectiveness of R-GAT in capturing important syntactic structures for sentiment analysis. The paper also discusses the influence of multiple aspects in one single sentence on sentiment prediction, showing that the aspects with high semantic similarity in a sentence may confuse the models. However, with R-GAT, both GAT and BERT can be improved across different ranges, showing that R-GAT can alleviate this problem to a certain extent.

Adversarial Training for Aspect-Based Sentiment Analysis with BERT, A. Karimi et al, 2021.

In the paper 'Adversarial Training for Aspect-Based Sentiment Analysis with BERT' the authors proposed a novel architecture that applies adversarial training to the fine-tuning

Category	Method	Restaurant		Laptop		Twitter	
		Accuracy	Macro-F1	Accuracy	Macro-F1	Accuracy	Macro-F1
Syn.	LSTM+SynATT	80.45	71.26	72.57	69.13	-	-
	AdaRNN	-	-	-	-	66.30	65.90
	PhraseRNN	66.20	59.32	-	-	-	-
	ASGCN	80.77	72.02	75.55	71.05	72.15	70.40
	CDT	82.30	74.02	77.19	72.99	74.66	73.66
	GAT	78.21	67.17	73.04	68.11	71.67	70.13
	TD-GAT	80.35	76.13	74.13	72.01	72.68	71.15
Att.	ATAE-LSTM	77.20	-	68.70	-	-	-
	IAN	78.60	-	72.10	-	-	-
	RAM	80.23	70.80	74.49	71.35	69.36	67.30
	MGAN	81.25	71.94	75.39	72.47	72.54	70.81
	LSTM	79.10	69.00	71.22	65.75	69.51	67.98
	BERT	85.62	78.28	77.58	72.38	75.28	74.11
Others	GCAE	77.28	-	69.14	-	-	-
	JCI	-	68.84	-	67.23	-	-
	TNET	80.69	71.27	76.54	71.75	74.90	73.60
Ours	R-GAT	83.30	76.08	77.42	73.76	75.57	73.82
Ours	R-GAT+BERT	**86.60**	**81.35**	**78.21**	**74.07**	**76.15**	**74.88**

Fig. 27. Overall performance of different methods on the three datasets.

process of BERT for aspect extraction and aspect sentimen classification tasks in sentiment analysis. Adversarial trainin; is a technique used to improve neural network performance b; introducing perturbed examples during training. The propose model outperformed general BERT and domain-specific post trained BERT in aspect-based sentiment analysis. The stud; also shows that the number of training epochs and dropou values significantly affect the model's performance.

Aspect-Based Sentiment Analysis involves two major tasks Aspect Extraction and Aspect Sentiment Classification. Aspec Extraction involves identifying and extracting all terms that point to aspects of a larger entity from a collection of review sentences. It is usually modeled as a sequence labeling task, with each word labeled as "Beginning," "Inside," or "Outside" of aspect terms. Aspect Extraction is typically performed using BERT architecture, which represents words as vectors and applies a fully connected layer to classify each word vector. Aspect Sentiment Classification involves classifying the sentiment towards each aspect in a review sentence as positive, negative, or neutral. Input sequences can have multiple aspects, which are addressed by preprocessing the input data for the model. Both Aspect Extraction and Aspect Sentiment Classification were subtasks of task 4 in the SemEval 2014 contest and have since been the focus of many studies.

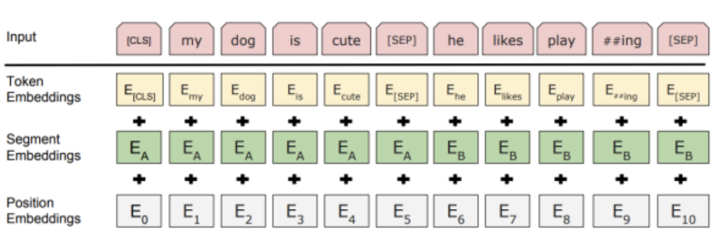

Fig. 28. BERT word embedding layer

The authors describe a proposed architecture for BERT adversarial training (BAT) to create adversarial examples from BERT embeddings using the gradient of the loss. The model consists of a BERT word embedding layer, a BERT encoder, a fully connected layer, and a loss function. The adversarial examples are created using a white-box method on the embedding level, where perturbations are added to the input embeddings to create new adversarial sentences. The adversarial loss is calculated using the BERT encoder, and the backpropagation algorithm is applied to the sum of both losses. The paper explains that BERT uses three different embeddings: token, segment, and position embeddings, which are computed by summing over each element. The BERT encoder is constructed using Transformer blocks and uses the Masked Language Model to preserve the information about the position of words in the sentence. The fully connected layer is used to classify the output embeddings of the BERT encoder into sentiment classes using cross-entropy loss.

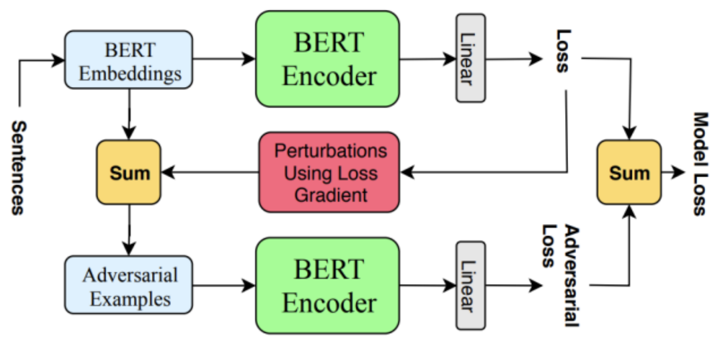

Fig. 29. The proposed architecture: BERT Adversarial Training (BAT).

The authors used benchmark datasets from SemEval 2014 task 4 and SemEval 2016 task 5 competitions to ensure consistency with previous works. They used a post-trained BERT model and experimented with different hyperparameters, such as training epochs and dropout values, to optimize the model's performance. They found that adversarial training improves the performance of the model for both tasks, with the best results achieved using a perturbation size of 0.3. The researchers also found that the results for the restaurant dataset were better than those for the laptop dataset, likely due to the selection of the validation set. The study's findings suggest that adversarial training can be an effective method for improving the performance of sentiment analysis models.

Aspect-based sentiment analysis using adaptive aspect-based lexicons, M. E. Mowlaei et al, 2020.

In the paper 'Aspect-based sentiment analysis using adaptive aspect-based lexicon' M. E. Mowlaei et al. Propose an Aspect-Based Frequency-Based Sentiment Analysis model which is an extension of the Frequency-Based Sentiment Analysis that allows it to analyze aspects within a sentence. It involves calculating the frequency of positive and negative aspects to which a word is most closely related in a sentence. The

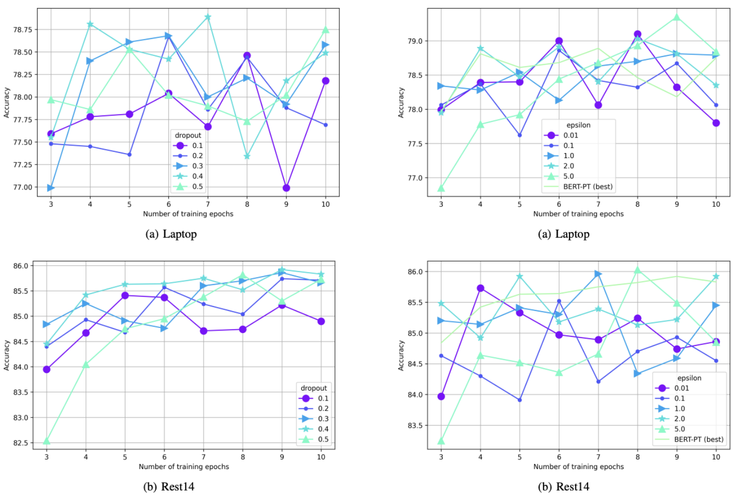

Fig. 30. Results on the impact of training epochs and dropout value in post-trained BERT for ASC task

algorithm also incorporates filtering of words by part-of-speech tags and lemmatization with the WordNet lemmatizer to increase the precision of the results. The effect of negations is also considered by incorporating a list of negations and reversing the sentiment of a word that is preceded by one of them. It can analyze multiple aspects within a single sentence.

The frequencies calculated are then normalized to account for class imbalance. Word scores are calculated based on these frequencies and are floating-point values in the range of -1 to +1. Words with scores greater than 0 are considered positive, while those with scores less than 0 are negative. Words with scores equal to 0 are neutral.

The technique is useful in sentiment analysis for social media, product reviews, and other domains where multiple aspects are discussed within a sentence. It is also useful for analyzing the sentiment of long documents or reviews. The algorithm is capable of providing accurate results and can be used to extract insights from large amounts of text data. The paper describes a genetic algorithm, which is used for aspect-based sentiment analysis. The algorithm takes as input a gene sequence and its fitness, along with the windows around the aspect terms. The mutate function selects candidate words from the windows and removes redundancies to form a set. The gene selection for alteration is then prioritized based on the candidate words. The crossover function takes two gene sequences and their respective fitness values as input. It locates genes in the donor sequence that are not identical to their respective genes in the parent sequence. Located genes are then extracted and replaced in a copy of the parent sequence. The function returns the current genes if there are improvements or tries again until it reaches a maximum attempt threshold.

The size of the parent pool and the maximum age of parents are the same as those in Mowlaei et al. (2018b). The algorithm starts with filling the pool with parents generated using the create function. Parents are selected using a roulette wheel selection scheme and modified using the mutate and crossover functions. The rate of mutation and crossover is determined dynamically based on their contributions to the improvements so far. The algorithm is useful for sentiment analysis and can help identify the sentiment of a sentence or paragraph towards specific aspects.

```
Input: dataset D<File>, lexicon ABALGA< Dictionary <String, Float>>,
lexicon BingLiu< Dictionary <String, Float>>,
lexicon MPQA< Dictionary <String, Float>>, lexicon SWN<SentiWordNet>
Output: precision<Float>, recall<Float>, F-measure<Float>

// Initialization
precisions ← Array<Float> (size: 10)
recalls ← Array<Float> (size: 10)
f_measures ← Array<Float> (size: 10)
// Extract terms, remove stopwords and lemmatize the terms
records, record_labels ← pre_process (D)
data_folds, label_folds ← 10_fold_split (data: records, labels: record_labels)
for i ← 1 to 10
    test_set ← data_folds [i]
    test_lebels ← label_folds [i]
    train_set ← concatenate (data_folds [ start : i], data_folds[ i+1 : end])
    train_labels ← concatenate (labels_folds [ start : i], labels_folds [ i+1 : end])
    ABFBSA<Dictionary<String, Float>> ← compute_abfbsa_word_scores (train_set)
    train_features ← extract_features (from: train_set, lexicons: [ABALGA, ABFBSA, MPQA, BingLiu, SWN])
    test_features ← extract_features (from: test_set, lexicons: [ABALGA, ABFBSA, MPQA, BingLiu, SWN])
    model ← Classifier ()
    train_model (classifier: model, data: train_features, labels: train_labels)
    predictions ← predict (classifier: model, test_features)
    p, r, f ← calculates_metrics (real_labels: test_labels, pred_labels: predictions, average: "Macro")
    precisions[i] ← p
    recalls[i] ← r
    f_measure[i] ← f
end for
return average (precisions), average (recalls), average (f_measures)
```

Fig. 31. Algorithm for the proposed framework

The article discusses the experimental setup and presents the results of a study on aspect-based sentiment analysis using a novel method. The study uses three datasets from Bing Liu's product review, referred to as benchmark datasets, to compare the technique with baseline methods. In addition, two datasets are used to evaluate different configurations of the method. The datasets are merged and shuffled, and a 10-fold cross-validation is used to assess the performance of the method. To prevent overfitting, the method is applied to the combination of four other datasets of domains related to the benchmark datasets, and the resulting lexicons with a fitness of at least 70% are stored.

The study reports that lexicons with a fitness of 93% were selected for the main classification task. The method extracted 3459 aspects of 1835 distinct words from datasets and the laptop training dataset of SemEval 2014. The results showed that using lexicons improved the performance of aspect-based sentiment analysis compared to baseline methods. The study also found that the method is stochastic in nature, and word scores from different runs on the same dataset may not be identical. Therefore, instead of using a single lexicon to extract features, three with a fitness of 93% were used for feature extraction. The study concludes that it is a promising method for aspect-based sentiment analysis and can be applied to other domains.

t-test results on F-measure values over 100 runs on benchmark datasets.

Base method	Compared to	t-statistics	p-value
ABFBSA	(Mowlaei et al., 2018a)	41.35	$<10^{-3}$
ABFBSA_bu	(Mowlaei et al., 2018a)	75.12	$<10^{-3}$
ABFBSA_bu	ABFBSA	36.71	$<10^{-3}$
ABALGA	(Mowlaei et al., 2018b)	−15.68	$<10^{-3}$
ABALGA_bu	(Mowlaei et al., 2018b)	20.54	$<10^{-3}$
ABALGA_bu	ABALGA	37.80	$<10^{-3}$
Comb_1	Comb_3	2.18	0.030
Comb_1	Comb_2	27.9	$<10^{-3}$
Proposed method	Comb_2	78.6	$<10^{-3}$
Proposed method	Comb_1	45.9	$<10^{-3}$

Fig. 32. t-test results on F-measure values over 100 runs on benchmark datasets

III. Code Implementation

Aspect Level Sentiment Classification with Deep Memory Network, D. Tang et al., 2016.

f1	training accuracy	validation accuracy
0.639	0.758	0.768

Adversarial Training for Aspect-Based Sentiment Analysis with BERT, A. Karimi et al., 2021.

f1	training accuracy	validation accuracy
0.726	0.856	0.834

Aspect-based Twitter Sentiment Classification, H. H. Lek et al., 2013.

```
print(aspect_scores[top_aspect])

# Step 4: Aspect classification
if aspect_scores[top_aspect] > 0:
  print(f"The {top_aspect} of the Starbucks Unicorn Frappuccino has a positive sentiment.")
else:
  print(f"The {top_aspect} of the Starbucks Unicorn Frappuccino has a negative sentiment.")

0.93
The taste of the Starbucks Unicorn Frappuccino has a positive sentiment.
```

Fig. 33. Aspect classification result

References

[1] G. Goldensohn, K. Hannan, and R. McDonald, Building a Sentiment Summarizer for Local Service Reviews, NLPIX, 2008.

[2] J. Zhu, M. Zhu, H. Wang, B. Tsou, Aspect-based sentence segmentation for sentiment summarization, CIKM, 2009.

[3] J. Zhu, H. Wang, B. Tsou, Multi-aspect opinion polling from textutual reviews, CIKM, 2009

[4] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, M. Stede, Lexicon-Based Methods for Sentiment Analysis, Computational Linguistics, 2011; 37 (2): 267-307

[5] : T. Na, J. Khoo, Christopher, Aspect-based sentiment analysis of movie reviews on discussion boards. J. Information Science, 2010 36. 823-848.

[6] : R. Varghese, M. Jayasree, Aspect based Sentiment Analysis using support vector machine classifier, ICACCI, 2013

[7] : H. H. Lek, D. C. C. Poo, Aspect-Based Twitter Sentiment Classification,IEEE, 2013.

[8] : W. Xue, T. Li, Aspect Based Sentiment Analysis with Gated Convolutional Networks, 2018

[9] : D. Tang, B. Qin, Aspect Level Sentiment Classification with Deep Memory Network, 2016

[10] : M. Hoang, O. A. Bihorac, J. Rouces, Aspect-Based Sentiment Analysis Using BERT, 2019

[11] : K. Wang, W. Shen, Y. Yang, X. Quan, R. Wang, Relational Graph Attention Network for Aspect-based Sentiment Analysis, 2020

[12] : M. E. Mowlaei, M. S. Abadeh, H. Keshavarz, Aspect-based sentiment analysis using adaptive aspect-based lexicons, 2020

[13] : A. Karimi, L Rossi, A. Prati, Adversarial Training for Aspect-Based Sentiment Analysis with BERT, 2021

[14] : A. Nazir, Y. Rao, L. Wu, L. Sun, Issues and Challenges of Aspect-based Sentiment Analysis, 2022

[15] : S. Kiritchenko, X. Zhu, C. Cherry, S. M. Mohammad, Detecting aspects and sentiment in customer reviews, 2014

[16] : Y. Kim, Convolutional Neural Networks for Sentence Classification, 2014

[17] : D. Tang, B. Qin, X. Feng, T. Liu, Effective LSTMs for Target-Dependent Sentiment Classification, 2016

[18] : D. Tang, B. Qin, and T. Liu, Aspect Level Sentiment Classification with Deep Memory Network, 2016

[19] : D. Ma, S. Li, X. Zhang, H. Wang, Interactive Attention Networks for Aspect-Level Sentiment Classification, 2017

[20] : P. Chen, Z. Sun, L. Bing, W. Yang Recurrent Attention Network on Memory for Aspect Sentiment Analysis, 2017

[21] : Hu M., Liu B., Mining and summarizing customer reviews, 2004