Indian Institute of Technology Guwahati,

Guwahati - 781039, Assam, India

CS-529: Topics and Tools in Social Media Data Mining

PROJECT REPORT

On

"A Unified Generative Framework for Aspect-Based Sentiment Analysis"

MTECH CSE 1ST YEAR

Group: ADMM

PROJECT MEMBERS:

Mayank Singh Parmar (214101027)

Devraj Gadhvi (214101017)

Amit Pawar (214101006)

Mohit Kumar (214101029)

1. Phase 1:

a. Brief introduction and background of research problem

[A unified generative framework for aspect-based sentiment analysis.]

Introduction:

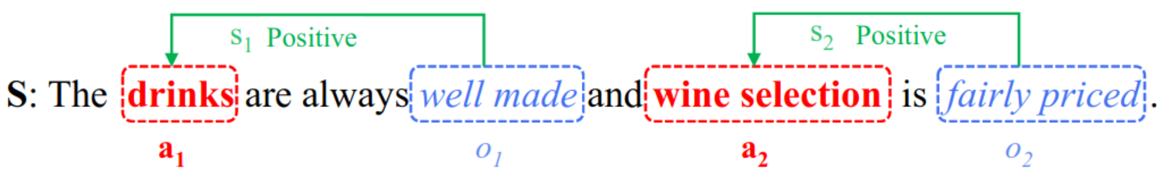
Aspect-based Sentiment Analysis (ABSA) aims to identify the aspect terms, their corresponding sentiment polarities, and the opinion terms.

Aspect-based Sentiment Analysis (ABSA) is the fine-grained Sentiment Analysis (SA) task, which aims to identify the aspect term (a), its corresponding sentiment polarity (s), and the opinion term (o).

There are two kinds of tasks: extraction task (extracting aspect and opinion)and classification task (predicting sentiment).

Some subtasks (AE, OE, AESC, Pair and Triplet) only take the text sentence as input. while the remaining subtasks (ALSC and AOE) take the text and a given aspect term as input.

There exist seven subtasks in ABSA.



Subtask	Input	Output	Task Type
Aspect Term Extraction(*AE*)	S	a_1, a_2	Extraction
Opinion Term Extraction(*OE*)	S	o_1, o_2	Extraction
Aspect-level Sentiment Classification(*ALSC*)	S + a_1	s_1	Classification
	S + a_2	s_2	
Aspect-oriented Opinion Extraction(*AOE*)	S + a_1	o_1	Extraction
	S + a_2	o_2	
Aspect Term Extraction and Sentiment Classification(*AESC*)	S	(a_1, s_1), (a_2, s_2)	Extraction & Classification
Pair Extraction(*Pair*)	S	(a_1, o_1), (a_2, o_2)	Extraction
Triplet Extraction(*Triplet*)	S	(a_1, o_1, s_1), (a_2, o_2, s_2)	Extraction & Classification

We formulate both the extraction task and classification task of ABSA into a unified index generation problem.
Paper uses a pre-training sequence-to-sequence model BART to solve all ABSA subtasks in an end to end framework.
We model the extraction and classification tasks as the pointer indexes and class indexes generation, respectively.

Background:

Some researches mainly focus on the single output subtasks. The AE, OE, ALSC and AOE subtasks only output one certain type from a, s or o while some researchers pay more attention and efforts to the subtasks with compound output

1. For AE subtask recent works explore sequence-to-sequence learning on AE subtask, which obtain promissing results especially with the pre-training language models.

2. Some papers incorporate the attentionmechanism into the LSTM-based neural network models to model relations of aspects and their contextual words for ALSC sub task.

3. For AESC one of paper follows pipeline method to solve this problem. Other works utilize unified tagging schema or multi-task learning to avoid the error-propagation problem . Span based AESC works are also proposed recently , which can tackle the sentiment inconsistency problem in the unified tagging schema.

4. To extract all (a,o) pair-wise relations from scratch. They propose a multi-task learning framework based on the span-based extraction method to handle this subtask.

5. Some methods apply the pipeline model to output the a, s, o from the inside sub-models separately. However, the pipeline process is not End-to-end.

   In conclusion, the existing methods can hardly solve all the subtasks by a unified framework without relying on the sub-models or changing the model structure to adapt to all ABSA subtasks.

b. Challenges in research problem:

The following divergences make it difficult to solve all subtasks in a unified Framework.

1. Input: Some subtasks (AE, OE, AESC, Pair and Triplet) only take the text sentence as input, while the remaining subtasks (ALSC and AOE) take the text and a given aspect term as Input.

2. Output: Some tasks (AE, OE, ALSC, AOE) only output a certain type from a, s or o, while the remaining tasks (AESC, Pair and Triplet) return compound output as the combination of a, s and o.

3. Task Type: There are two kinds of tasks:
    extraction task (extracting aspect and opinion) and
    classification task (predicting sentiment).

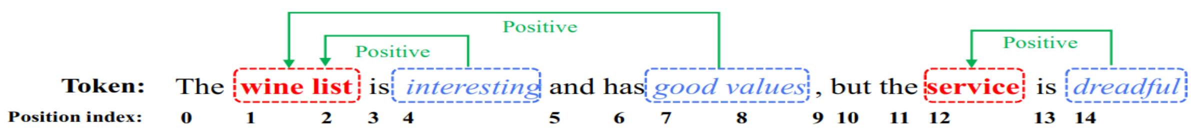
Thus, previous works only focus on the subset of these subtasks. However, the paper solves the whole ABSA subtasks in a unified framework using the BART model.

c. Proposed Model:

We model the extraction and classification tasks as the pointer indexes and class indexes generation, respectively.

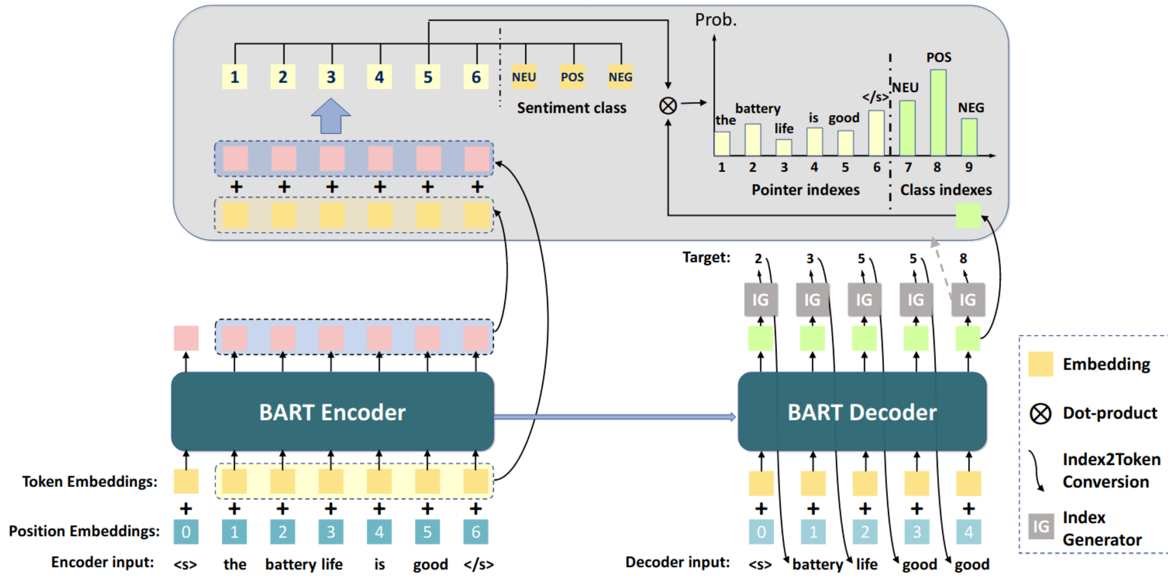
Task formulation: We use a, s, o, to represent the aspect term, sentiment polarity, and opinion term, respectively.

.



Subtask	Target Sequence
AE	1, 2, 12, 12, </s>
OE	4, 4, 7, 8, 14, 14, </s>
ALSC	1, 2 , POS, </s>
	12, 12, POS, </s>
AOE	1, 2, 4, 4, 7, 8, </s>
	12, 12, 14, 14, </s>
AESC	1, 2, POS, 12, 12, NEG, </s>
Pair	1, 2, 4, 4, 1, 2, 7, 8, 12, 12, 14, 14, </s>
Triplet	1, 2, 4, 4, POS, 1, 2, 7, 8, POS, 12, 12, 14, 14, POS, </s>

The overall architecture of the framework



This shows an example generation process for the Triplet subtask where the source is “<s>the battery life is good </s>” and the target is “2 3 5 5 8 6”(Only partial decoder sequence is shown where the 6 (</s>) should be the next generation index).

The “**Index2Token Conversion**” converts the index to tokens. Specifically, the pointer index will be converted to its corresponding token in the source text, and the class index will be converted to corresponding class tokens. Embedding vectors in yellow boxes are retrieved from the same embedding matrix. We use different position embeddings in the source and target for better generation performance.

About BART:

BART (Bidirectional and Auto-Regressive Transformers) is a strong sequence-to-sequence pretrained model for Natural Language Generation (NLG).

BART is a denoising autoencoder composed of several transformer encoder and decoder layers.

BART- Base model contains a 6-layer encoder and 6-layer decoder, which makes it similar in number of parameters with the BERT-Base model.

BART is pretrained on denoising tasks where the input sentence is noised by some methods, such as masking and permutation.

The encoder takes the noised sentence as input, and the decoder will restore the original sentence in an autoregressive manner.

2. Phase 2:

(a) Limitations of proposed model in research paper.

The proposed model of the research paper is limited to finding only the aspect terms ,opinion terms ,polarity and sentiment classification and does not tell about the stance of the sentence with the target.

Given paper is about sentiment analysis but there is no established relation between sentiment and stance of a given sentence with respect to the given target.

Sentiment analysis classifies text as positive, negative or else objective, so it can be thought of as a text classification task. Text classification has many classes as there are many topics but sentiment analysis has only three classes. However, there are many factors that make sentiment analysis difficult compared to traditional text classification.

Tweet	Target	Sentiment	Stance Label
@rimmedlarry Actually, the tag was made by feminists so they can narcissistically post selfies to prove they're not ugly.	Feminist Movement	Negative	Against
SO EXCITING! Meaningful climate change action is on the way!	Climate Change is a Real Concern	Positive	Favour
When the last tree is cut down, the last fish eaten & the last stream poisoned, you will realize that you cannot eat money.	Climate Change is a Real Concern	Negative	Favor
dear lord thank u for all ofur blessings forgive my sins lord give me strength and energy for this busy day ahead.	Atheism Sentiment	Positive	Against

(b) Objective of CS-529 phase 2 project:

To detect the stance of a given sentence with respect to a target

(c) Explanation of intuitions behind objective.

Nowadays, social media platforms constitute a major component of an individual's social interaction. These platforms are considered robust information dissemination tools to express opinions and share views. People rely on these tools as the main source of news to connect with the world and get instant updates . They are very beneficial because they allow individuals to explore various aspects of emerging topics, express their own points of view, get instant feedback, and explore the views held by the public. The huge dependency of users on these platforms as their main source of communication allows researchers to study different aspects of online human behavior, including the public stance toward various social and political aspects.

Stance is defined as the expression of the speaker's standpoint and judgment toward a given proposition. Stance detection plays a major role in analytical studies measuring public opinion on social media, particularly on political and social issues. The nature of these issues is usually controversial, wherein people express opposing opinions toward differentiable points.

Social issues such as abortion, climate change, and feminism have been heavily used as target topics for stance detection on social media. Similarly, political topics, such as referendums and elections, have always been hot topics that have been used in stance detection to study public opinion .

The stance has been used in various research as a means to link linguistic forms and social identities that have the capability to better understand the background of people with polarized stances . Consequently, the early work on stance detection emerged in analyzing political debates on online forums.

The majority of work on stance detection has targeted the detection of the stance toward a given subject expressed in a given text. Thus, the work on stance detection can be categorized into two main types, namely, detecting expressed views versus predicting unexpressed views. In the first type, the objective is to classify a user's post and infer the current stance to be in favor of or against a given subject.

The work on stance detection can also be categorized based on the topic of the target of analysis, that is, it can be one specific target, multiple related targets, or a claim in a news article. Most existing designs of stance detection classifiers work to identify the user's stance toward a specific topic. Sometimes, the classifier is built to detect the stance toward multiple related targets.

Applications of Stance Detection

1) Opinion Mining.
    1) People's opinion on the new policy by Govt.
    2) Analyzing the speech of members in different houses.
    3) Analyzing employee opinion of an organization towards a new policy

2) Analyzing Debates.
    1) Online Debate.
    2) Debate in parliament.
    3) Analyzing the comment section of a Facebook post. (Survey for News agencies)
    4) Analyzing conversation between groups of people.

3) Feedback System.

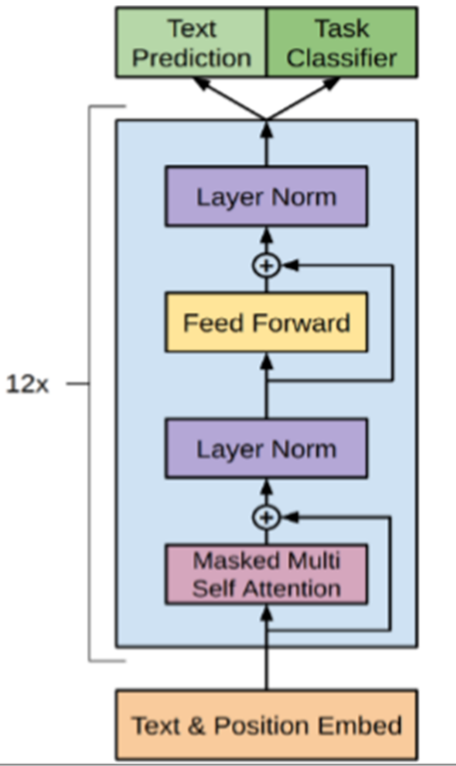
1) Feedback given by a customer on Amazon, Flipkart etc.
2) Analyzing chat boxes.
3) Analysis of Public opinion for a policy which is implemented by Govt.
4) People experience new products launched by a company.

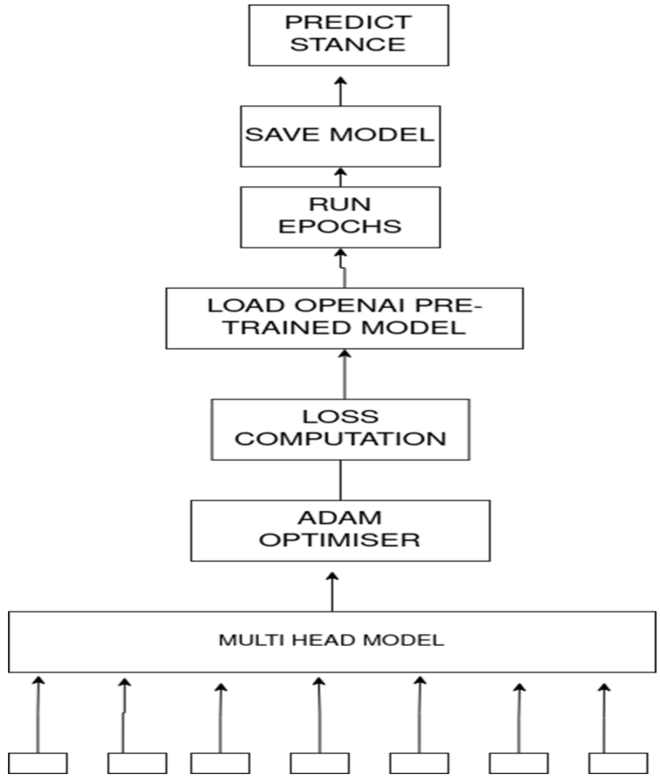
4) Analysis of social media.
1) Tweet by twitter users.
2) Post by Facebook users.
3) Post on Instagram and different other social media platforms.

5) Security.
1. Fake news Detection.
2. Rumors Detection.

(d) Supporting experimental setup.

- We trained a 12-layer decoder-only transformer with masked self-attention head(768 dimensional states and 12 attention heads)
- We used a cross entropy loss function used to optimize the model during training.
- We used the Adam optimization scheme with a max learning rate of 2.5e-5
- We used a byte pair encoding (BPE) vocabulary with 40,000 merges and residual, embedding, and attention dropouts with a rate of 0.1 for regularization.
- and residual, embedding, and attention dropouts with a rate of 0.1 for regularization
- Each minibatch was capable of sampling a maximum of 512 contiguous tokens and the number of batches was 8.
- Dataset is of SemEval 2016 for testing and training.
- The transformer uses spaCy for tokenization.we use wrapper for a byte-pair encoded tokenizer, In order to use spaCy's English tokenizer,

- 

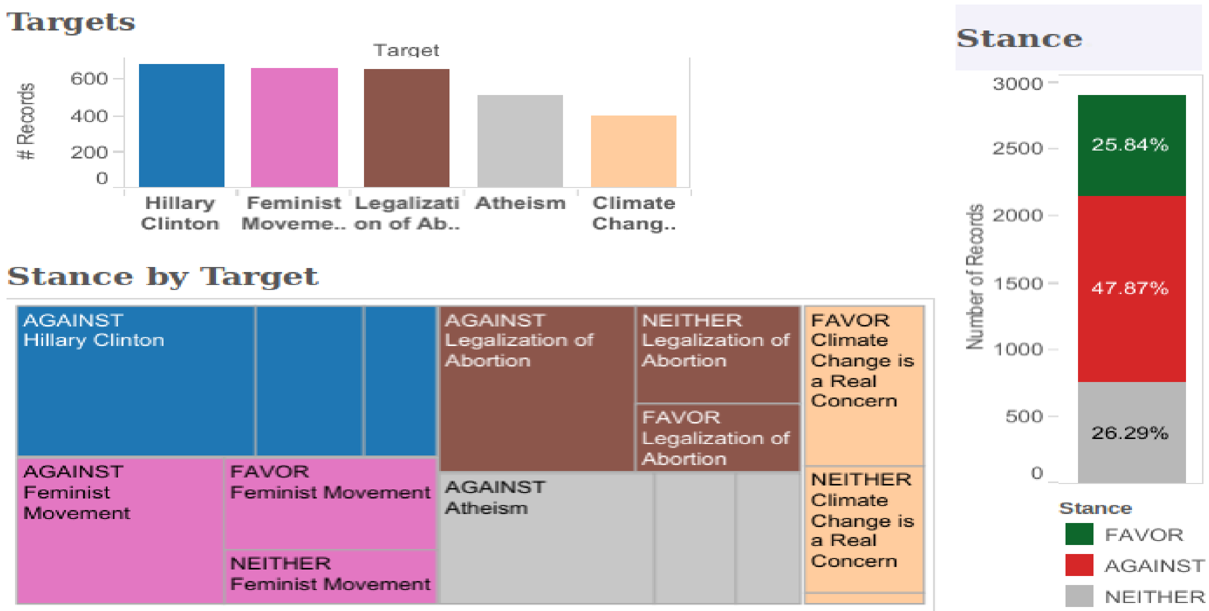


(e) Produced results over data.

Dataset: SemEval Task 6

The total number of Tweets (in the training set) available for this task is roughly 2700, which amounts to roughly 500-600 Tweets per topic. Thus, this can be considered a small dataset.

	Target	Tweet	Stance
669	Feminist Movement	@Maisie_Williams is our hero with her #LikeAGi...	FAVOR
670	Feminist Movement	Rather be an "ugly" feminist then be these sad...	FAVOR
671	Feminist Movement	iamNovaah: RT ChrzOC: Bitches be running wild....	AGAINST
672	Feminist Movement	@angerelle you disagree that people should str...	AGAINST
673	Feminist Movement	#Rapeculture is basically a FABLE. It has almo...	AGAINST



In the result we are getting accuracy score of 54% using SemEval testfile

(f)Results in terms of accuracy, class-wise F-measure.

```
Logging
1 281 35.659 36.169 50.62 49.73
running epoch 1
Logging
2 562 22.452 23.972 46.18 47.25
running epoch 2
Logging
3 843 34.305 42.338 55.06 52.22
running epoch 3
Logging
4 1124 17.200 32.332 71.05 57.02
running epoch 4
Logging
5 1405 15.928 42.220 77.98 58.97
running epoch 5
Logging
6 1686 10.830 47.213 86.32 60.75
running epoch 6
Logging
7 1967 10.033 57.467 88.45 60.75
              precision    recall  f1-score   support

     AGAINST       0.65      0.74      0.69       715
       FAVOR       0.69      0.03      0.06       304
        NONE       0.31      0.58      0.41       230

    accuracy                           0.54      1249
   macro avg       0.55      0.45      0.39      1249
weighted avg       0.60      0.54      0.48      1249
```

Numbers after the Logging are:

Epoch number ,number of updates , training _cost, validation_cost, training _accuracy, validation accuracy

(g) Observation from experimental results.

Training one single classifier for all topics at once (running the training loop on the entire training dataset) it was noticed that when trying to train the transformer on just a single topic (which had < 500 training samples), there was significant over-fitting where the validation accuracy dropped to well below 60%. This could be because the transformer has a high-dimensional embedding layer (768 dimensions) that requires a sufficient amount of training data to avoid over-fitting.

Overall, fine-tuning the dropout, the weight of the language modeling objective and changing some of the other default arguments (such as random seed) did little to improve the macro F-score and accuracy.

(h) Future work .

Our future works includes:

- Increase the accuracy score in less iterations.
- Use other models like BART , BERT and compare their obtained accuracies .
- Try to obtain a relation between sentiment and stance
- Use this stance and sentiment analysis to do analysis of social media.