Aspect Based Sentiment Analysis

Anonymous ACL submission

Abstract

This report presents our work done in the lieu of course project component in Natural Language Processing(Monsoon 2021) Course.Our project topic is Aspect Based Sentiment Analysis.It is a sub-task of sentiment classification problem, in which sentiment is predicted corresponding to some particular term in the input sentence.We will discuss different sub tasks that are involved at different instances of the problem and also show our experimentation results to the models we picked for implementation.

1 Introduction

1.1 Problem Statement

The goal is to identify the aspects of given target entities and the sentiment expressed towards each aspect.

1.2 Motivation

The proliferation of social media and e-Commerce platforms such as Facebook, Amazon has provided a place for customers to provide feedback to businesses based on their happiness. Customers' reviews are a worldwide trusted source of authentic material for other users. Sentiment Analysis is used in process of analysing tweets, blog posts, or comments, categorising the users' perspectives in order to determine consumer sentiment about the product. It enables organisations to get a true sense of how their customers "feel" about their brand and handle massive volumes of data in a timely and cost-effective manner. Brands can listen intently to their consumers and adjust goods and services to match their demands by automatically assessing client input, from survey replies to social media chats. Aspect-based sentiment analysis (ABSA) divides data into aspects and determines the sentiment associated with each. Customer feedback associates specific attitudes with different characteristics of a product or service. The seller needs to know whether specific characteristics or features are being mentioned positively, neutrally, or negatively. ABSA would be able to detect that the phrase indicates a negative judgement about the feature battery life in this text: "The battery life of this smartphone is very bad.

1.3 Online Deployment Link

The project is deployed on local host.

1.4 Project Pipeline Summary

The pipeline of our project Aspect Based Sentiment Analysis widely comprises of two sub tasks. First is extraction of Aspect Terms in the input sentence, Second is Sentiment classification of the sentence corresponding to extracted aspect terms. General pipeline of the project includes pre-processing of input sentences , converting the sentences into its embedding representation, extracting aspect terms, feeding extracted aspect terms context along with input sentence for its sentiment classification and finally we show (aspect,sentiment) pair for each aspect in the sentence.

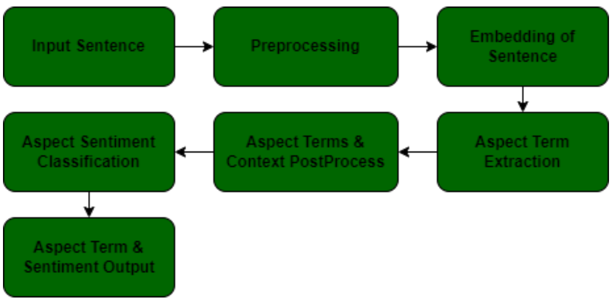

Figure 1: General Pipeline for ABSA.

2 Related Work

In this section, we survey some representative studies in the aspect-based sentiment analysis (ABSA). ABSA aims to predict the sentiment towards given terms or categories. Two major tasks in ABSA are ATE(Aspect Term Extraction): extracting aspect terms from given sentence and APC(Aspect Polarity Classification): Classify polarity of aspects identified in ATE. The section below explains about the work done in both these tasks separately.

2.1 Aspect Term Extraction

To identify aspects, most early ABSA studies used a set of filters on high-frequency noun phrases. A noun, adjective, verb, or adverb can all be used to indicate an aspect. To filter out non-aspects, several filtering approaches are employed to frequent nouns. These models had a major drawback that they need to be manually modified according to new dataset which was practically impossible.

Following these, sequential models were used to extract aspect terms using HMM and CRF.CRF-based models for fine-grain sentiment analysis are proposed by (Shariaty and Moghaddam, 2011). By learning the parameters from the data automatically, the sequential technique overcomes the constraints of language rule methods. Irrespective of good performance, they require labelled data to train which was a limitation in many cases.

After these classical machine learning methods, deep learning methods came into picture to detect aspect terms.Deep learning methods use distributed vectors to automatically learn latent characteristics and has outperformed several machine learning approaches on similar tasks. Each review in ABSA may include more than one aspect, and hence more than one class, a basic supervised model cannot categorise each review into distinct classes. Several deep learning architectures have been built to tackle this difficulty.(Ma et al.) propose interactive attention networks, which train attention in both phrases as well as aspects simultaneously and build representations for each individually. The challenge was viewed as a BIO sequence labelling by (Liu et al., 2015).

Based on RNNs and word embedding, a wide class of discriminative models are there. (Wang et al., 2016a) created a Recursive Neural Conditional Random Fields model by combining CRF with a Recursive Neural Network. Clearly deep learning models with attention and transformers outperformed ATE task compared to machine learning methods.

2.2 Aspect Polarity Classification

Aspect Polarity Classification is a classification task where based on given aspect, model needs to classify the polarity to be positive, negative or neutral. One important challenge in this task is to model the relationship among aspect terms and sentences. (Jiang et al.) used rich features to model this relationship but these methods are labour intensive. As neural networks are capable of learning features without feature engineering, Deep Learning came into picture for APC task.

(Huang and Carley, 2019) used PF-CNN and PG-CNN are two new neural units to combine aspect information into CNN. One shortcoming of the traditional CNN is its lack of understanding of information derived from aspect words. By parameterizing filters using aspect terms, PG-CNN solves this flaw. The number of general characteristics that should flow to the final classification is controlled by a parameterized gate in PG-CNN. For ABSA, the LSTM with attention mechanism has been a preferred structure but the training takes a long time. Because they can mine semantics in contextual windows, CNNs are effective and can extract meaningful local patterns.

Recently (Tang et al., 2016) used Target-Dependent LSTM (TD-LSTM) and Target-Connection LSTM (TC-LSTM) to identify relationship among aspects and words however they were not able to capture that effectively. With introduction to attention mechanisms, aspect-based sentiment analysis with attention models (Wang et al., 2016b) proved to be more effective for this classification tasks. (Gu et al.) also takes position of words with respect to aspect along with attention to classify sentiment.

Current techniques ((Wu and Ong)) on APC used Bert based embeddings and models to understand deep semantics within the text and capture accurate relationship between aspect and sentence words.

3 Methodology

Aspect Based Sentiment Analysis(ABSA) mainly consist of two major task : Aspect Term Extraction(ATE) and Aspect Polarity Classification (APC) or Aspect Based Sentiment Classification (ABSC) .

3.1 Aspect Term Extraction

Aspect Term Extraction is a sentence labelling problem in which for each word/token in the input sentence we provide a label to it using 'B,I,O' labelling scheme.In this scheme B tag is given to a token which is starting of an aspect term,I tag is given to tokens representing intermediate aspect terms and O is for rest of tokens which are non aspect terms.

3.1.1 Bi-directional LSTM based Model

By intuition we can understand that for solving this problem context of neighbour words are very useful.Hence a possible model can involve Recurrent Neural Networks for learning it.We know vanilla RNNs suffers from vanishing gradient problems, hence researchers prefers to use its improved version LSTM units for developing time series models.Since for a given token we know both its past and future tokens, we can utilize Bidirectional LSTMs which can learn more contextual concepts. So we use this sequential BiLSTMs as our first classifier.
We have tested different combinations of embeddings e.g. pre-trained GLove.300d, trained from scratch Fastext embedding and Bert embeddings(further information about combinations used is given in Experiments and Analysis sections).After this layer we used Dense Layer of size 3 with softmax activation function for providing confidence outputs for each label.

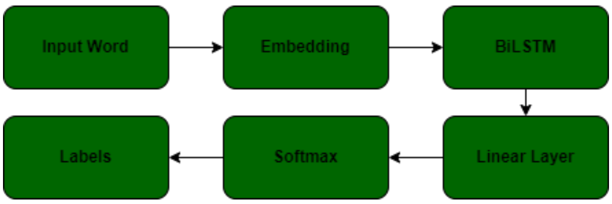

Figure 2: LSTM Based Pipeline for ATE

3.1.2 CNN Based Model

CNNs are widely used for image related deep learning problems but recently it is also getting adopted in NLP domain .The reason behind using using CNNs are they are simple models which are quick to train, it can capture useful features from inputs and propagate it to subsequent CNN layers for capturing higher level features.Taking reference from (Xu et al., 2018) we have also implemented a CNN based Aspect Term Extraction Classifier.As given in reference, it uses two different types of embeddings as input, one embedding is designed from a large corpus general context embedding model, here we used GLove.300d, we call them global embeddings, along with this we also train a domain specific embedding which contains semantic information of tokens that is specific to the domain.In this model architecture we have concatenation of general and domain embedding as input to Conv1D layer. Filter size for first layer is kept to 3 and 5 for subsequent layer.Zero Padding is used for corners. there are 3 more CNN layers are used followed by a softmax Dense layer of size 3, which will predict the probability of output labels. While experimenting we have also combined BERT embeddings which improves the performance.

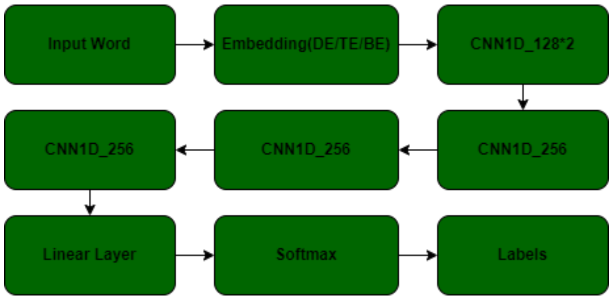

Figure 3: CNN based Model Pipeline for ATE

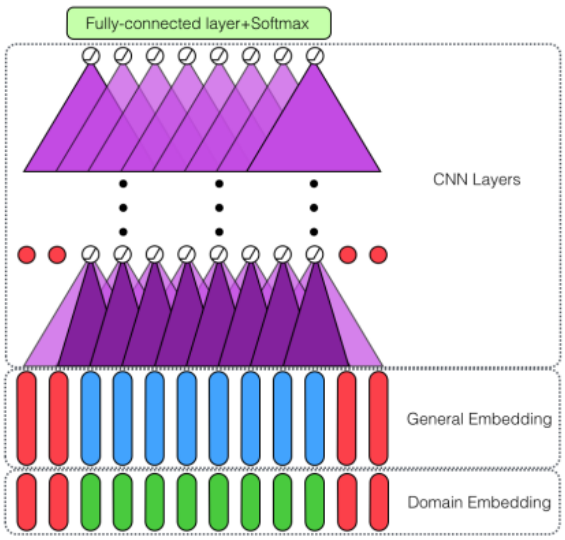

Figure 4: Visual pipeline for CNN based ATE using Double Embedding ref (Xu et al., 2018)

3.2 Aspect Polarity Classification

APC is a fine-grained classification task: Given input a sentence and labeled aspect terms in the sentence classify polarity with respect to each aspect.

In such fine-grained classification, it is important to capture details with respect to different aspects for example the sentiment polarity of ***"staffs are not that friendly, but the taste covers all."*** will be positive with respect to food but negative when considering the aspect service. The same sentence can have different polarities with respect to different aspects.

Datasets used for this task are SemEval 2014 Restaurant and Laptop datasets. Data is preprocessed and sentiments are put into three major classes: Positive, Negative, and Neutral. If a sentence contains two different aspects it is replicated into two different instances one concerning each aspect. In total 4 different features are required namely: text, aspect_text, position_information, and sentiment. The text contains tokenized words (indices) padded to the maximum length sentence in the dataset. The aspect_text contains aspect words (indices) to the maximum length sentence in the dataset. The position_information contains information about the closeness of each word from the aspect term in form of indices where the aspect term is given index 0 and index of words to the left of aspect term and the right of aspect term start from 1 and so on. Sentiment contains sentiment of each instance.

3.2.1 Standard LSTM based Model

The Baseline model for Aspect Polarity Classification is a standard Long Short Term Memory Network(LSTM). The words in sentences are glove encoded and passed to the LSTM network and classification is performed using the hidden state at the last timestamp. The LSTM model used contains only one single layer and is unidirectional and classifies polarity among three classes positive, negative and neutral.

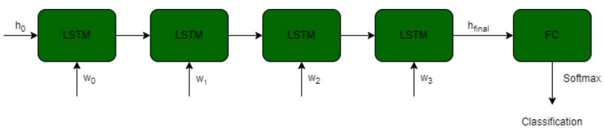

Figure 5: Standard LSTM Model of APC

3.2.2 LSTM based Model with Aspect Embedding : AE-LSTM

The baseline LSTM model mentioned above does not consider any aspect information for classification. As we know aspect information is important for polarity classification and the same sentence can have two different polarities concerning different aspects we developed another LSTM model considering aspect information. Model-1 is standard LSTM with input as word glove vectors appended along with aspect glove vectors at each time step. Final Classification is performed using the last hidden representation.

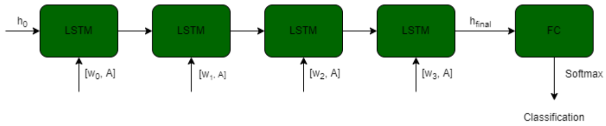

Figure 6: Aspect Embedding-LSTM Model of APC

3.2.3 Position Aware Bi-directional Attention Network : PBAN

The above Model-1 simply appends the target vector at each timestamp but does not closely capture how important each word is with respect to a given aspect term this is where Attention comes into the picture. Also, it does not capture the importance of position information. Generally, words closer to aspect term are more important for example : **"It's a perfect place to have amazing Indian food."** Here the aspect term is Indian food and the word just next to it "amazing" has a greater role in deciding the polarity of the sentence rather than words farther away. Based on this analysis we developed Model-2 Position Aware Bi-directional Attention Network (PBAN).

The final model uses two different Gated Recurrent Units (GRU). The input to GRU-1 is aspect embedding and the input to GRU-2 is concatenated vector of position information randomly initialized embedding and word embedding. The hidden representation from both GRUs are used to calculate Bi-directional attention weights that is "aspect term to position-aware attention" and "position-aware sentence attention to aspect term", using bi-directional attention helps to focus on important words both in sentence and aspect with respect to each other. The output from attention is multiplied with weights in the final layer and classification is performed.

4 Experiments

We have different experiments while implementing different models through the project.For ATE task , for both models we have tested out different embedding combinations of Glove(General), Fastext(Domain) and BERT(Contextual),It is also shown in the results that triple embedding performs

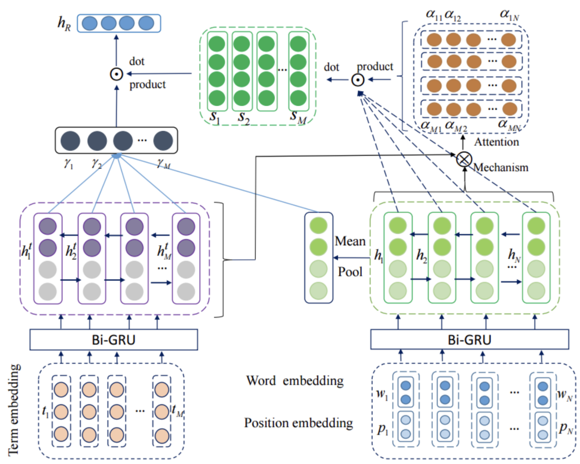

Figure 7: Position Aware Sentence Attention Pipeline for APC figure taken from (Wang et al., 2016b)

very well in sequence labelling.We have also included dropout layers, to introduce regularization in the models , however it did not helped us in terms of performance.We have tried and tuned hyper-parametrs like number of epochs, batch size etc. and chosen the hyper-parametrs which gives better results.

In APC tasks we have tried over three different models : LSTM, AE-LSTM and PBAN using glove embedding for text. For each model we have performed hyper parameter tuning over learning rate, number of epochs, early stopping, dropout, number of layers in lstm, bi-directional vs uni-directional lstm and different optimizers with different initialization techniques. For APC task, PBAN model performs the best.

5 Result and Analysis

We are presenting our analysis on both the sub-tasks separately.For ATE, different versions of models we created are annotated as :

a)**DE-CNN**: General and Domain Embedding based CNN model

b)**TE-CNN**: General,Domain and Bert Embedding based CNN model

c)**BE-CNN**: Only BERT embedding based CNN model.

d)**DE-LSTM**: General and Domain Embedding based LSTM model

e)**TE-LSTM**: General,Domainand BERT Embedding based LSTM model

f)**BE-LSTM**: Only Embedding based LSTM model

Since the nature of dataset for ATE task is imbalanced we will use F1 score as evaluation metric :

Evaluation for ATE on Laptop Dataset			
Model Type	TE	DE	BE
LSTM	0.69	0.55	0.45
CNN	**0.75**	0.71	0.55

Evaluation for ATE on Restaurant Dataset			
Model Type	TE	DE	BE
LSTM	0.81	0.83	0.62
CNN	0.82	**0.831**	0.62

Triple embedding based CNN outperforms for Laptop Dataset while Double embedding based CNN works best with Restaurant Dataset.

Evaluation for APC on Laptop Dataset	
Model Type	Score
LSTM	0.32
AE-LSTM	0.57
PBAN	**0.67**

Evaluation for APC on Restaurant Dataset	
Model Type	Score
LSTM	0.39
AE-LSTM	0.52
PBAN	**0.61**

For APC task PBAN model performs the best over both the datasets because it uses bi-directional attention to capture the semantic relationship between aspects and words in sentence.

Evaluation for ATE and APC on Laptop		
ATE Model	APC Model	Score
LSTM with Double	LSTM	0.37
CNN with Triple	PBAN	**0.55**

Evaluation for ATE and APC on Restaraunt		
ATE Model	APC Model	Score
LSTM with Double	LSTM	0.31
CNN with Triple	PBAN	**0.62**

According to the above analysis we can conclude that CNN with Triple Embedding for ATE and PBAN for APC performs the best. We concluded that embeddings play a crucial role in model training and with rich embeddings result can be improved. Also understanding the relation ship between aspect and words is important for APC task which is done through Bi-Directional Attention here.

References

Shuqin Gu, Lipeng Zhang, Yuexian Hou, and Yin Song. A Position-aware Bidirectional Attention Network for Aspect-level Sentiment Analysis. page 11.

Binxuan Huang and Kathleen M. Carley. 2019. Parameterized Convolutional Neural Networks for Aspect Level Sentiment Classification. *arXiv:1909.06276 [cs]*.

Long Jiang, Mo Yu, Ming Zhou, Xiaohua Liu, and Tiejun Zhao. Target-dependent Twitter Sentiment Classification. page 10.

Pengfei Liu, Shafiq Joty, and Helen Meng. 2015. Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1433–1443, Lisbon, Portugal. Association for Computational Linguistics.

Yukun Ma, Haiyun Peng, and Erik Cambria. Targeted Aspect-Based Sentiment Analysis via Embedding Commonsense Knowledge into an Attentive LSTM. page 8.

Shabnam Shariaty and Samaneh Moghaddam. 2011. Fine-Grained Opinion Mining Using Conditional Random Fields. In *2011 IEEE 11th International Conference on Data Mining Workshops*, pages 109–114, Vancouver, BC, Canada. IEEE.

Duyu Tang, Bing Qin, Xiaocheng Feng, and Ting Liu. 2016. Effective LSTMs for Target-Dependent Sentiment Classification. *arXiv:1512.01100 [cs]*.

Wenya Wang, Sinno Jialin Pan, Daniel Dahlmeier, and Xiaokui Xiao. 2016a. Recursive Neural Conditional Random Fields for Aspect-based Sentiment Analysis. *arXiv:1603.06679 [cs]*.

Yequan Wang, Minlie Huang, xiaoyan zhu, and Li Zhao. 2016b. Attention-based LSTM for Aspect-level Sentiment Classification. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 606–615, Austin, Texas. Association for Computational Linguistics.

Zhengxuan Wu and Desmond C Ong. Context-Guided BERT for Targeted Aspect-Based Sentiment Analysis. page 9.

Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu. 2018. Double embeddings and CNN-based sequence labeling for aspect extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 592–598, Melbourne, Australia. Association for Computational Linguistics.