SMAI PROJECT REPORT
TEAM 28
ASPECT BASED SENTIMENT ANALYSIS WITH GATED CNN

Team Members -

1. Venneti Sri Satya Vinay-2020101066
2. Janapati Sai Akash-2020101046
3. Rahul Mallavarapu-2020101053
4. Nagireddy Dhanush-2020101061

Base Paper used:

https://arxiv.org/pdf/1805.07043.pdf

Base Problem:

Consumers and service providers can benefit from the sentiment analysis and opinion mining of user-generated reviews. It is suggested that fine-grained aspect based sentiment analysis (ABSA) be used to better understand reviews rather than traditional sentiment analysis, which just predicts the overall sentiment polarity. ABSA is divided into two subtasks:

- Aspect-Category Sentiment Analysis (ACSA)
- Aspect-Term Sentiment Analysis (ATSA)

Our Approach:

- Most of the existing work employ Long Short-term memory (**LSTM**) and attention mechanisms for predicting sentiment polarity in a sentence which required more training time.

- This solution proposes a model using Convolutional Neural networks (**CNN**) which are more fast and accurate.
- Gated Tanh-ReLU Units (**GTRU**) units proposed in this solution can selectively output the sentiment features according to the given aspect or entity.
- The computations proposed in this model can be parallelised as CNNs do not have time dependency like LSTM layers.

ACSA, ATSA Example:

Consider the sentence

"*Average to good Thai food, but terrible delivery.*"

- Here ATSA would ask the sentiment polarity of **specific terms** in the text which could be multi-word phrase or a single word like "*Thai Food*".
- ACSA would ask the sentiment polarity toward a **particular aspect** like *service* even though it does not appear in the sentence. Here the aspect categories are pre-defined.

Architecture:

- A new model for ACSA and ATSA is proposed namely, Gated Convolutional network with Aspect Embedding (GCAE).
- The CNN model consists of an **embedding layer**, a **one-dimension convolutional layer** and a **max-pooling layer**.
- The embedding layer takes the indices $w_i \in \{1, 2, \ldots, V\}$ of the input words and outputs the corresponding embedding vectors $v_i \in R^D$ where D is dimension of embedding vectors. It is usually initialised with pre-trained embeddings such as GloVe.
- The one-dimensional convolutional layer has multiple kernels each extracting a n-gram at various granularities.
- $X = [v_1, v_2, \ldots, v_L]$ where L is the length of the sentence. Here X is the input sentence after passing through embedding layer.
- A convolutional filter $W_c \in R^{D \times k}$ maps k words in the receptive field to a single feature c. So we get a set of new features $[c_1, c_2, \ldots, c_L]$.
- $c_i = f(X_{i:i+K} * W_c + b_c)$
- This is what happens in a traditional CNN.
- Here we put Gated Tanh-ReLU Units (GTRU) with aspect embedding are

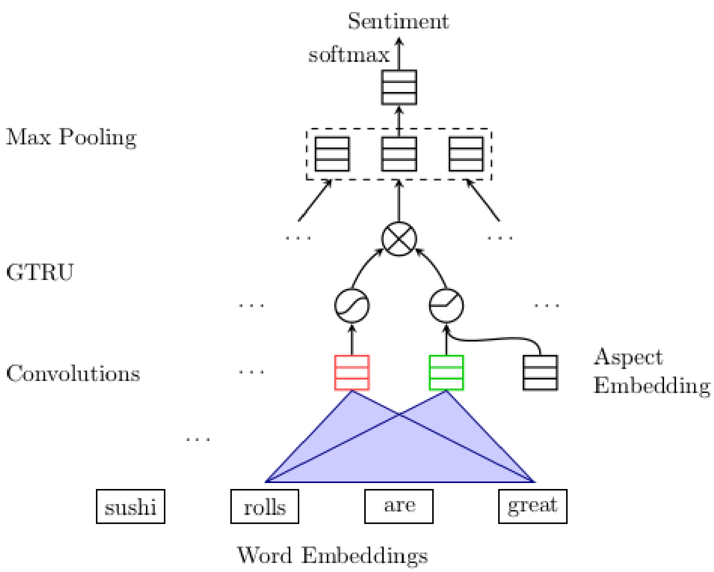
connected to two convolutional neurons at each position t. Specifically, we compute the features c_i as

$$
\begin{aligned}
a_i &= \text{relu}(\mathbf{X}_{i:i+k} * \mathbf{W}_a + \mathbf{V}_a \boldsymbol{v}_a + b_a) \\
s_i &= \tanh(\mathbf{X}_{i:i+k} * \mathbf{W}_s + b_s) \\
c_i &= s_i \times a_i \quad ,
\end{aligned}
$$

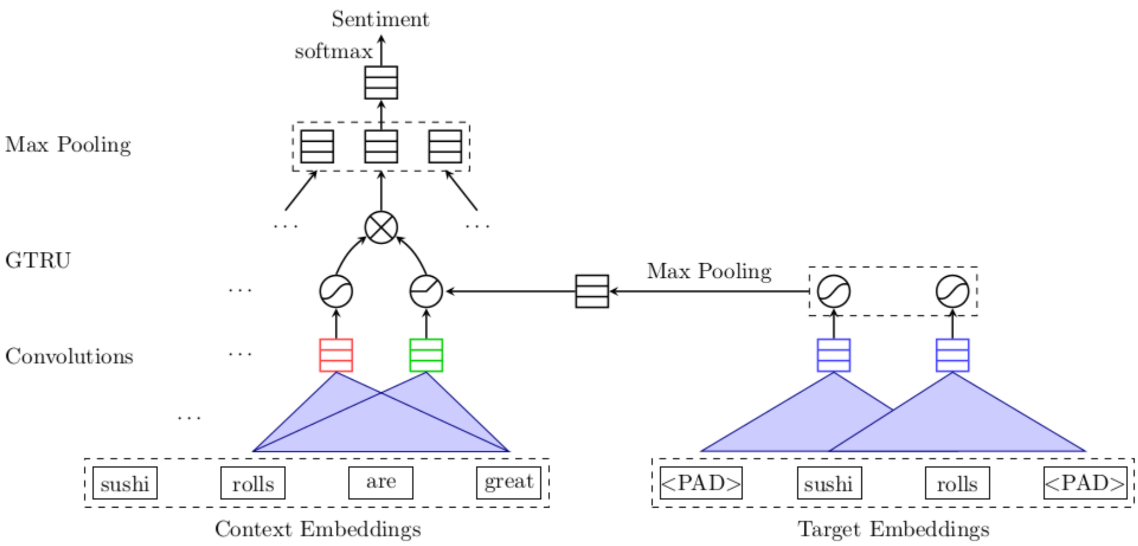
- Here v_a is the embedding vector of given aspect category in ACSA.
- So the first two equations are similar to traditional CNN but have additional aspect category v_a with ReLU activation function.
- s_i and a_i are responsible for generating sentiment features and aspect features respectively. We then take product of both to get c_i.
- Then we do max-over-time layer that generates a fixed size vector of size e $\in$ R^{d_k} which keeps most significant features of the sentence.
- We then use softmax function to predict sentiment polarity y_{pred}.
- We train the model by minimising the cross-entropy loss between ground truth y and predicted y, y_{pred}. Cross-entropy is given by

$$\mathcal{L} = -\sum_i \sum_j y_i^j \log \hat{y}_i^j$$

GCAE on ACSA:



GCAE on ATSA:



- The purpose of the ATSA task is to foretell the aspect term's sentiment polarity.
- A small convolutional layer on aspect terms is added to extend Gated Convolutional network with Aspect Embedding (GCAE).
- While in aspect-term sentiment analysis (ATSA), this information is provided by a small convolutional neural network (CNN) on aspect terms, in aspect-category sentiment analysis (ACSA), it comes from a single aspect word and controls the flow of sentiment features in the Gated Tanh-ReLU Units (GTRU).
- The supplementary CNN maintains the capability of parallel computation while extracting the key features from many words.

Gating Mechanisms:

- The flow of sentiment data to the pooling layer is controlled by the Gated Tanh-ReLU Units (GTRU).
- Positive inputs to the ReLU gate have no upper bound, while negative inputs must be exactly zero. As a result, it can produce a similarity score based on how relevant the provided aspect information and the aspect feature are.
- The sentiment features would be blocked outright, if this score is 0, otherwise, their amplitude would be enhanced in accordance.
- The sentiment features that are insignificant for the entire sentence are further eliminated by the max-over-time pooling.

- On text classification problems, GTRU has been more effective than Gated Tanh Units (GTU) and Gated Linear Units (GLU).

Datasets:

- Datasets from the SemEval Workshops have been used
- For ACSA, datasets of SemEval Task 4 2014 Have been used, which contains 5 aspects and 4 polarities.
- A larger dataset has been created by eliminating conflict class from the above and combining reviews from 2015, 2016 to create a larger dataset.
- We use SemEval 2014 Task 4 restaurant and laptop reviews for the ATSA task.
- Also small datasets with mixed sentiments based on aspect are separated to create a "hard" dataset, to evaluate the performance better.We use SemEval 2014 Task 4 restaurant and laptop reviews for the ATSA task.

Comparision:

Some benchmark models have been used to compare the performance of GCAE:

- **NRC-Canada:** The top method in SemEval 2014 Task 4 for ACSA and ATSA task. SVM is trained with extensive feature engineering.
- **CNN:** Very strong baseline for sentiment classification, but cannot capture aspect information directly.
- **TD-LSTM :** Creates target-dependent representation for sentiment prediction using two LSTM networks to model the target's prior and following contexts.
- **ATAE-LSTM:** An attention-based LSTM for ACSA task, embeds given aspect embedding with each word embedding as the input of LSTM.
- **IAN:** For ATSA, uses LSTM with attention mechanisms.
- **RAM:** Similar to the previous, but with multiple attention layers.
- **GCN:** Gated CNN, in which GTRU does not have the aspect embedding as an additional input.

Results:

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**

Models	Restaurant		Laptop	
	Test	Hard Test	Test	Hard Test
SVM*	77.13	-	63.61	-
SVM + lexicons*	**80.16**	-	**70.49**	-
TD-LSTM	73.44±1.17	56.48±2.46	62.23±0.92	46.11±1.89
ATAE-LSTM	73.74±3.01	50.98±2.27	64.38±4.52	40.39±1.30
IAN	76.34±0.27	55.16±1.97	68.49±0.57	44.51±0.48
RAM	76.97±0.64	55.85±1.60	68.48±0.85	45.37±2.03
GCAE	**77.28±0.32**	**56.73±0.56**	**69.14±0.32**	**47.06±2.45**

Why GCAE works:

- The extracted information from sentences must be divided into aspect information and sentiment information. The two types of information must be conveyed simultaneously via the context vectors produced by the LSTM. The similarity scoring function also produces attention scores for the complete context vector.
- In order to control the sentiment information flow according to the provided aspect information at each dimension of the context vectors, our model first incorporates GTRU.
- Second, in order to unravel aspect and sentiment information, GCAE generates two context vectors, one for each features aspect and sentiment.

Training time:

- The architecture can be trained parallelly, as most of convolution operations, gates and pooling layers are independent.
- This reduces the training time drastically as the existing models like LSTMs process the data sequentially.

Model	ATSA
ATAE	25.28
IAN	82.87
RAM	64.16
TD-LSTM	19.39
GCAE	3.33

Conclusion:

- Our model implements Convolution Networks and Gated Mechanisms for the task of ACSA and ATSA, which were previously implemented using LSTM and Attention Mechanisms.
- But the older implementations had larger training times and didn't support parallel computations, thus making them unable to make use of advances in hardware that enable parallel computations.
- Our newer model has both better accuracy and supports parallel computation.
- We lately also compared our neural network with various other relevant neural networks that were frequently used for the same purpose.

Our Implementation

Papers referred:

- "*A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification*"
  https://arxiv.org/abs/1510.03820
- "*Do Convolutional Networks need to be Deep for Text Classification ?*"
  https://arxiv.org/abs/1707.04108
- "*Convolutional Neural Networks for Sentence Classification*"
  https://arxiv.org/pdf/1408.5882.pdf

What we implemented:

- We implemented ACSA for SemEval ACSA-restaurant large dataset.We achieved an accuracy of 99% on this dataset in the training phase and 76% in the testing phase.The base paper achieved an accuracy of 85% on the same dataset.
- We implemented ATSA for SemEval ATSA-restaurant dataset.We achieved an accuracy of 98% on this dataset in the training phase and 72% in the testing phase.The base paper achieved an accuracy of 77% on the same dataset.

Github Repo:

https://github.com/srisatyavinay/SMAI-Project-Team28