SMAI

A hierarchical model of reviews for aspect-based sentiment analysis

Team Details

Team name: huh lmao

Team members: Anushka Jain (2021111008), Tanish Taneja (2021112011)

Introduction

This project is inspired by the research paper titled "A Hierarchical Model of Reviews for Aspect-based Sentiment Analysis" by Rudger and Breslin. The paper challenges the conventional approach of independently classifying sentences within customer reviews and proposes a hierarchical bidirectional LSTM model to capture the intricate relationships among sentences. This project aims to replicate and further investigate the paper's findings and compare them with our implementations of baseline models mentioned in the paper, potentially contributing to the advancement of aspect-based sentiment analysis.

Motivation behind the project

This project is driven by the compelling findings of the referenced research paper, which demonstrates the potential for significant improvements in aspect-based sentiment analysis through the use of a hierarchical model. The project's motivation lies in the desire to leverage these innovative techniques to enhance our understanding of customer reviews and sentiment analysis, ultimately benefiting various domains reliant on accurate opinion mining without the need for resource-intensive hand-engineered features.

Deliverables

The project requires the following deliverables from us:

Baseline models

Implementing the baseline models—CNN and LSTM—mentioned in the paper and analyzing their performance on our datasets. To ascertain that the hierarchical nature of the proposed model is the deciding factor, it is compared against a sentence-level convolutional neural network and a sentence-level Bi-LSTM, which is identical to the first layer of our model. According to the results section, these models are also used in the provided paper as baseline models.

Proposed model

Implementing the proposed model, a hierarchical bidirectional LSTM model for sentiment analysis. Stacking a Bi-LSTM on the review level on top of sentence-level Bi-LSTMs yields the hierarchical bidirectional LSTM (H-LSTM). This model is then trained and tested on our databases.

Analysis

Analyzing the merits and demerits of the baseline and proposed models thoroughly.

Timeline

Deadline	Milestone	Tasks	
7 Oct	CHECKPOINT 1	Reading the Paper	Making Deliverables Doc
15 Oct		Data Preprocessing	Working with CNNs and LSTMs in PyTorch
23 Oct		Implementing Baseline CNN Model	
30 Oct	CHECKPOINT 2	Implementing Baseline Bi-LSTM Model	Comparing Baselines
10 Nov		Implementing Hierarchical Bi-LSTM Model as discussed in paper	
20 Nov		Training and Fine Tuning the Model	Comparison with other Models
27 Nov	CHECKPOINT 3	Analysis of Results and Conclusions	
30 Nov	FINAL DEADLINE	Submission	

Datasets

For the project, we will use the same datasets used in the research paper to find the results obtained by the models we implemented. We consider datasets in five domains (restaurants, hotels, laptops, phones, cameras) and eight languages (English, Spanish, French, Russian, Dutch, Turkish, Arabic, and Chinese) from the recent SemEval-2016 Aspect-based Sentiment Analysis task, using the provided train/test splits. There are 11 domain-language datasets containing 300-400 reviews with 1250-6000 sentences. Each sentence is annotated with none, one, or multiple domain-specific aspects and a sentiment value for each aspect.