Aspect Based Sentiment Analysis using Deep Learning

CS585, UMass Amherst, Fall 2016

Satya Narayan Shukla, Utkarsh Srivastava

Introduction

- Sentiment Analysis is the process of classifying opinions expressed in text as positive, negative or neutral.
- In the context of customer reviews, an overall rating for a product may not be entirely representative.
- An example could be 2 reviews of a single product having same overall rating but conflicting aspect opinion.
- We aim to extend this approach to predict the sentiment of each aspect present in the review.
- Although, deep learning methods have exhibited great power in sentiment analysis tasks, the functionalities of these have not been entirely exploited for this problem.

Methodology

- **Pre-processing Step**: Represent less frequently occurring aspects in train set as 'Other' and include a new aspect tag 'None'.
- **Input Representation**: A sentence matrix where each row represents a word as a 300-dimensional vector (WE) trained on Google/Amazon corpus.
- **Additional Features (AF)**:

-- 7 POS-tag features (noun, verb, adjective, adverb, preposition, conjunction and other) extracted using nltk tagger
-- A feature for similarity with given aspect words using WordNet
-- 2-features for cosine similarity with given '+ve' and '-ve' words
-- A feature for identifying negation words
-- Re-scaling word-embedding with info from above defined features
-- Removing words not present in the word embedding and stop words

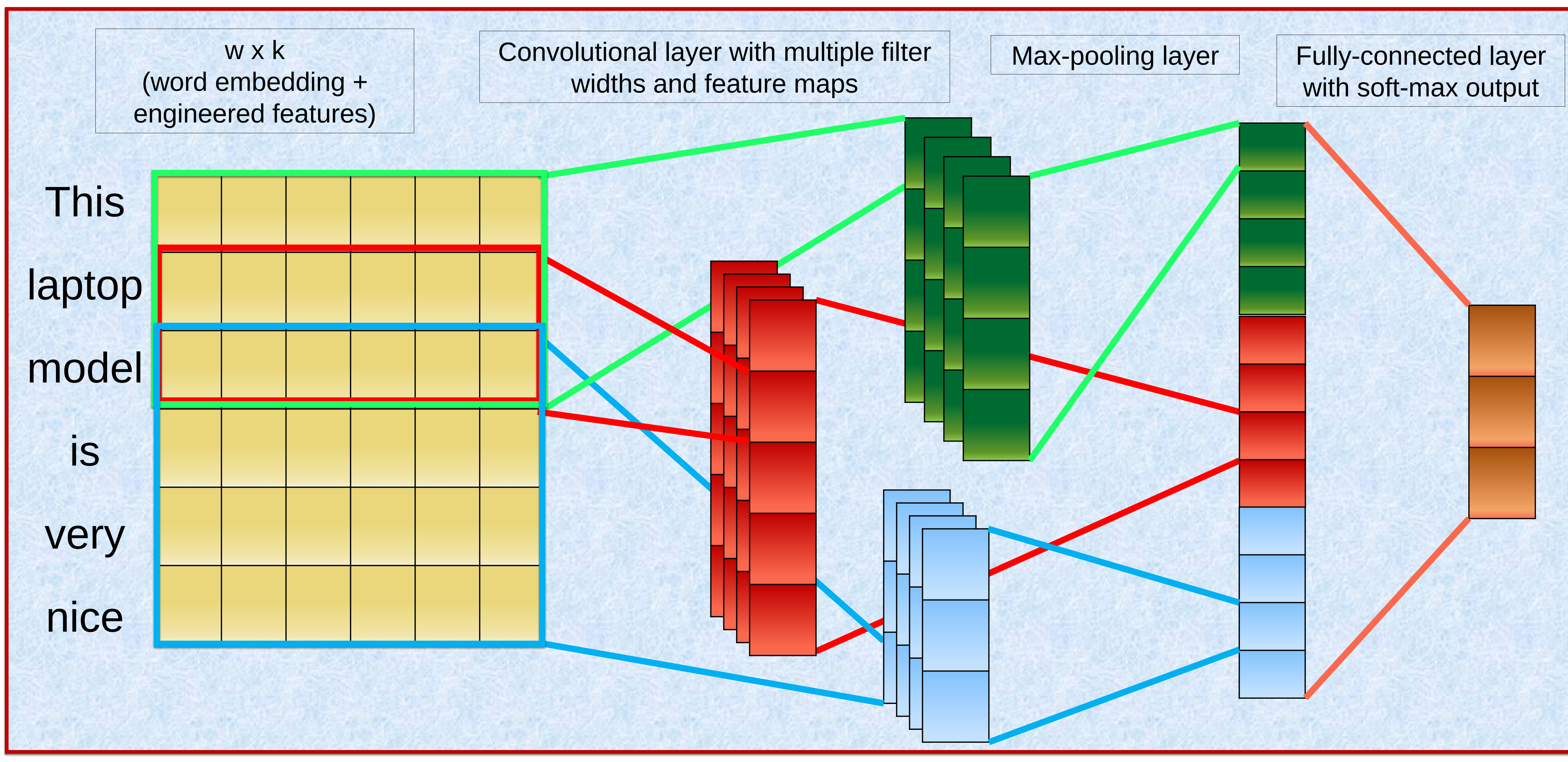
Dataset

- Dataset : **SemEval '15 – Task 12**
- Each input sentence comprises of Aspect(s) (as in Entity#Attribute) and associated Polarity.


```
- <Review rid="339">
  - <sentences>
    - <sentence id="339:0">
        <text>Plain and simple, it(laptop) runs great and loads fast.</text>
      - <Opinions>
          <Opinion polarity="positive" category="LAPTOP#OPERATION_PERFORMANCE"/>
        </Opinions>
      </sentence>
    - <sentence id="339:1">
        <text>Easy to carry, can be taken anywhere, can be hooked up to printers,headsets.</text>
      - <Opinions>
          <Opinion polarity="positive" category="LAPTOP#PORTABILITY"/>
          <Opinion polarity="positive" category="LAPTOP#CONNECTIVITY"/>
        </Opinions>
      </sentence>
    - <sentence id="339:2" OutOfScope="TRUE">
        <text>Love that it doesn't take up space like a regular computer.</text>
      </sentence>
    </sentences>
  </Review>
```


Domain	Train Data	Test Data	Aspect Types in Train Data	Aspect Types in Test Data
Laptop	1739	761	81	58
Restaurant	1315	685	13	12

Domain	Aspect labels in Raw Data	Aspect labels after pre-processing	Aspect#Polarity Labels
Laptop	81	25	73
Restaurant	13	14	40



Network

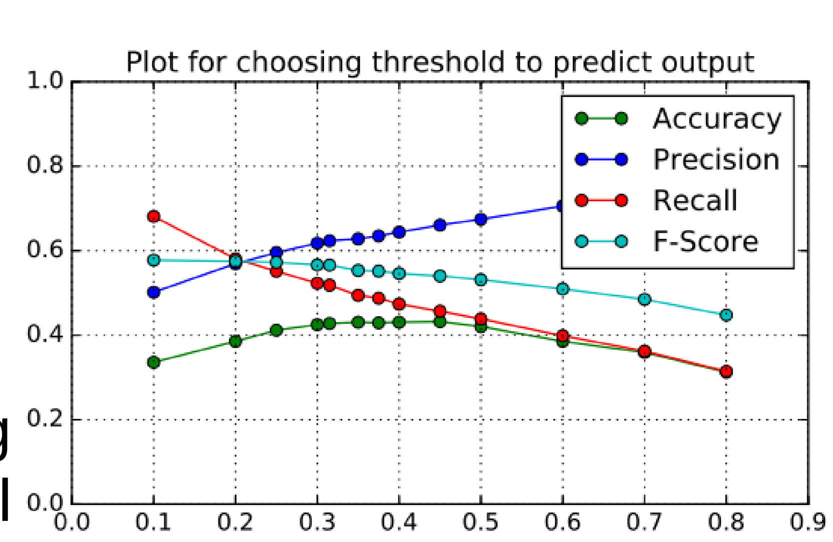
- Input Layer (WE)
- Convolutional Layer with 100 feature maps of filter size 2 and 3. Output computed using tanh.
- Max-pool layer to extract maximum value.
- Dense Layer (800 nodes) with 50 % dropout and L2 norm of the weights. Output computed using ReLU unit.
- Soft-max output layer

Trained in Batch size = 50

Results

Domain	Framework	Aspect			Aspect + Polarity		
		Precision	Recall	F-score	Precision	Recall	F-score
Restaurant	Google WE	**65.0 %**	**53.5 %**	**58.7 %**	**53.9 %**	**39.5 %**	**45.6 %**
	Amazon WE	45.1 %	36.8 %	40.5 %	37.1 %	21.6 %	27.3 %
	Google WE + AF	64.9 %	51.7 %	57.5 %	52.1 %	37.2 %	43.4 %
	Amazon WE + AF	49.0 %	37.8 %	42.7 %	36.7 %	28.0 %	31.8 %
Laptop	Google WE	**59.6 %**	**47.9 %**	**53.1 %**	**50.7 %**	**33.7 %**	**40.5 %**
	Google WE + AF	57.6 %	37.4 %	45.4 %	42.6 %	30.2 %	35.3 %

- Hyper-parameters (filter size, filter count, max-pool ridge) selected based on accuracy results from multiple runs on a random train-validate split.
- Substituting Amazon embedding lowered the result values in all cases.



Comparison

- Baseline : SVM (trained with linear kernel)
- Comparison for 'Aspect Category Identification' based on F1-Score.

Domain	Our Model	Team Rank @ SemEval '15					Baseline
		1	2	3	4	5	
Restaurant	58.7 %	**62.7 %**	61.9 %	57.2 %	57.1 %	54.1 %	51.3 %
Laptop	**53.1 %**	50.9 %	50.0 %	49.6 %	49.0 %	46.5 %	48.1 %

- As can be seen from the above table, our results are comparable to the models yielding the best results for SemEval '15 – Task 12.

Challenges

- Gaining accuracy for low frequency count aspects present in training data is difficult.
- Finding a general threshold value for output layer to handle multiple aspects in a single sentence.
- Words not present in WE model may lead to incomplete model representation.