Aspect-Based Sentiment Analysis using Deep Neural Networks and Transfer Learning

Technische Universität München

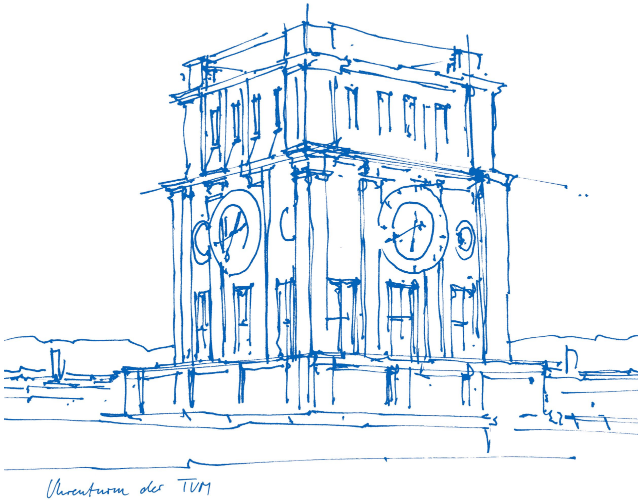
Fakultät für Informatik

Presented by : Sumit Dugar

Supervisor : PD Dr. Georg Groh

Advisor : Gerhard Hagerer

Date : 11.04.2019



Outline

1. Introduction
    a. Project Introduction and Motivation
    b. Problem Statement and Proposed Solution
2. Details
    a. Proposed Method
        i. Input Transformation
        ii. Architecture
    b. Transfer Learning
        i. Progressive Neural Network (PNN)
        ii. Single Unified Network (SUN)
3. Future work & Conclusion

Introduction

Project Introduction and Motivation

- What is Aspect-Based Sentiment Analysis?

  Food was good but the service was bad.

  food#quality - positive   service#general - negative

- What motivates us for pursuing this project?
  - A step towards general purpose AI.
  - A lot of scope for business applications.

Problem Statement and Proposed Solutions

- Effective incorporation of contextual information.
    - Hierarchical BiLSTMs.
    - Attention Mechanism.

- Huge number of output classes.
    - Proposed Input transformation.

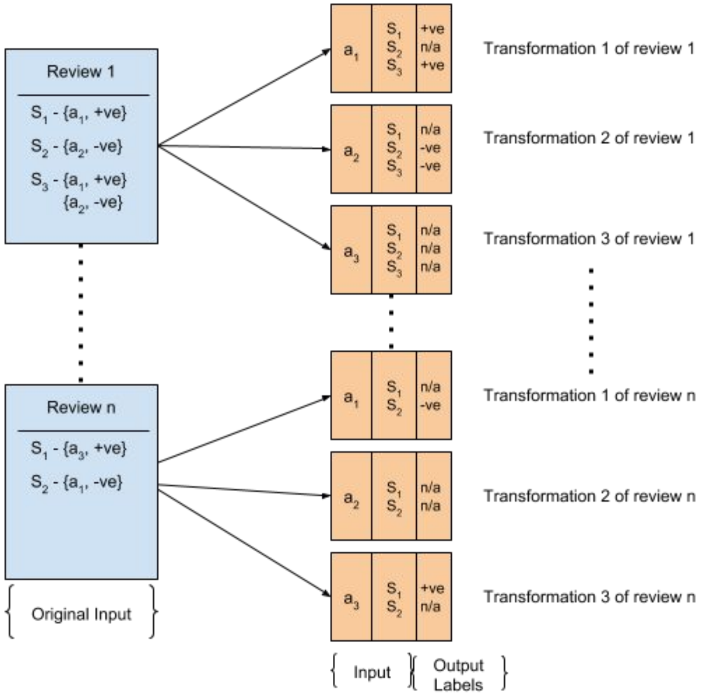
Problem Statement and Proposed Solutions

- Limited and unbalanced training data.
  - Data augmentation.
  - Proposed Input transformation.

- Ability to transfer knowledge from source to target domain.
  - Progressive Neural Networks (PNNs).
  - Single Unified Network (SUN).

Details

Proposed Input Transformation

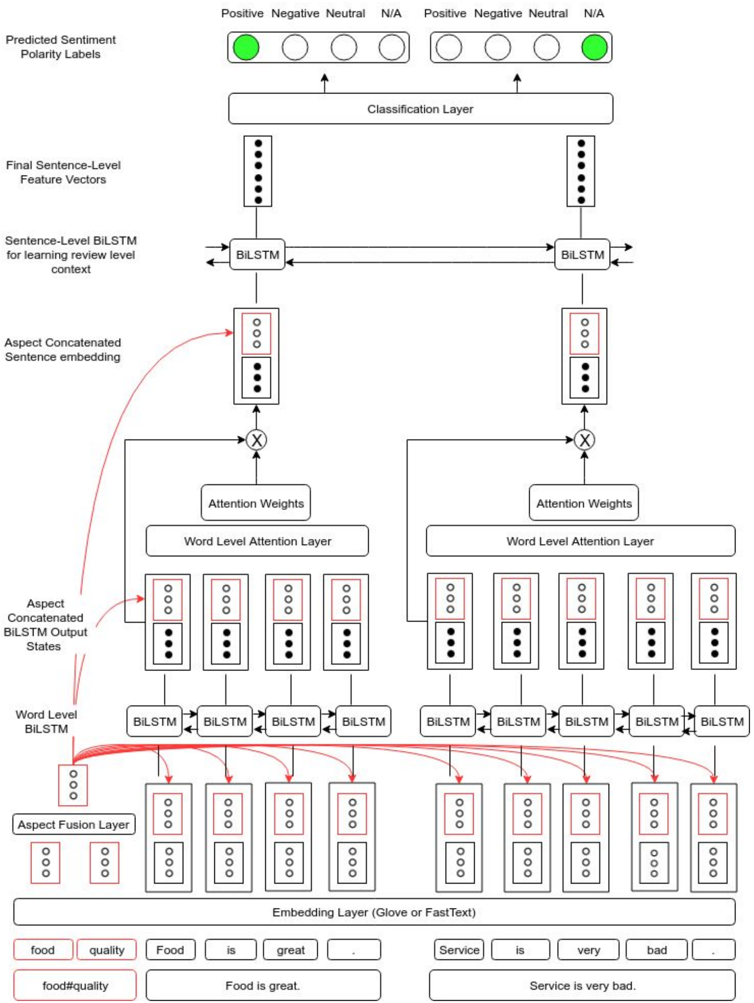


Some advantages of transforming the data in this format are :

- Fixed number of output classes.
- Graceful handling of multi-label scenarios.
  - No thresholding.
  - No sentence duplication.
- Data augmentation

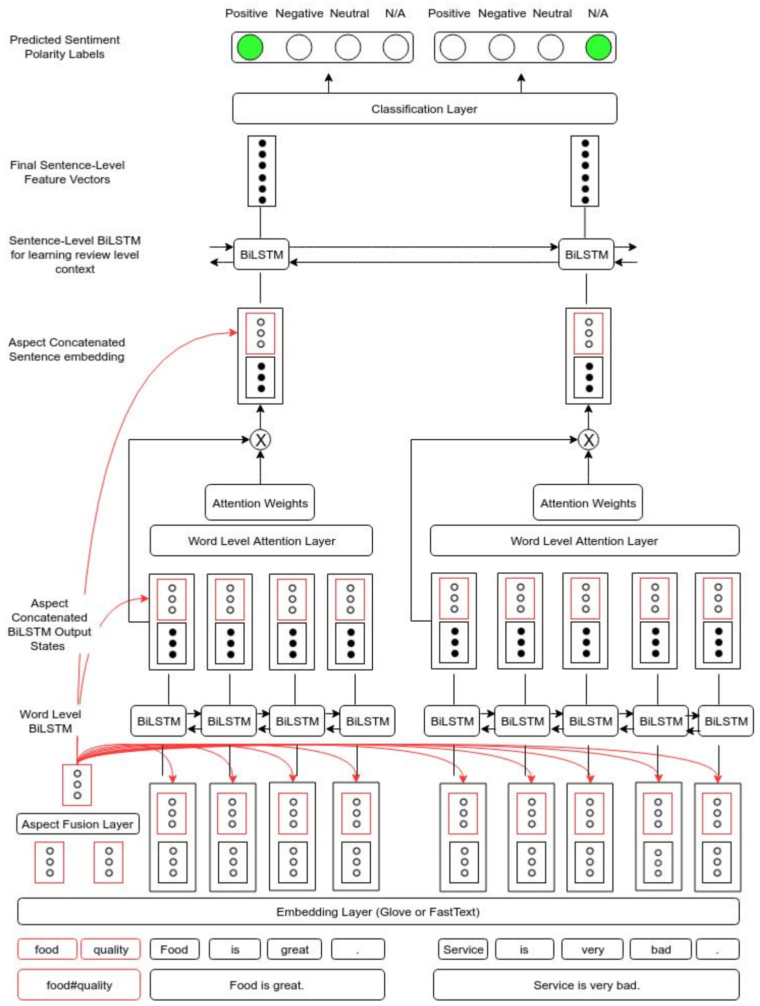
Proposed Architecture

- A joint approach for simultaneous aspect and sentiment detection.
- Aspect words are fused together for generating fixed sized aspect representation.



Proposed Architecture

- Hierarchical nature of the model helps to capture the context by combining information from neighboring words and sentences.
- Sentence representations are generated by applying attention on word vectors.



Experiments and Results - Word Embeddings

- Considerable improvement over baseline results.
- SemEval-2016 Restaurant worked best with GloVe and SemEval-2016 Laptops worked best with fastText.
- Fine-tuning of embeddings resulted in worse results.

SemEval-2016 Restaurant

Variant	Macro F1 Score					Micro F1 Score
	average	*positive*	*negative*	*neutral*	*n/a*	
Baseline	0.47	0.551	0.378	0.000	0.962	0.337
PT-G	0.57	0.688	0.418	0.205	0.969	0.454
PT-F	0.52	0.580	0.323	0.222	0.951	0.354
PT-E	0.54	0.686	0.478	0.000	0.970	0.455
FT-G	0.53	0.632	0.307	0.188	0.967	0.379
FT-F	0.48	0.566	0.329	0.047	0.957	0.331

PT-G=Pre-Trained-GloVe, PT-F=Pre-Trained-fastText, PT-E=Pre-Trained-ELMo, FT-G=Fine-Tuned-GloVe, FT-F=Fine-Tuned-fastText

SemEval-2016 Laptops

Variant	Macro F1 Score					Micro F1 Score
	average	*positive*	*negative*	*neutral*	*n/a*	
PT-G + CWL + DA	0.42	0.393	0.208	0.054	0.993	0.220
PT-F + CWL + DA	0.44	0.426	0.255	0.068	0.993	0.246

PT-G=Pre-Trained-GloVe, CWL=Class Weighted Loss, DA=Data augmentation, PT-F=Pre-Trained-fastText

Experiments and Results - Combating Data Imbalance

- By using Class weighted Loss (CWL).
- With the help of Data Augmentation (DA).
- CWL & DA improved performance across both the embeddings.

SemEval-2016 Restaurant

Aspects (Data Count)	PT-G	PT-G & CWL	PT-G & CWL & DA	PT-F	PT-F & CWL	PT-F & CWL & DA
food-style-options (137)	0.586	0.481	0.487	0.315	0.313	0.359
food-quality (849)	0.769	0.804	0.823	0.808	0.786	0.802
drinks-prices (20)	0.000	0.000	0.000	0.000	0.181	0.571
food-prices (90)	0.482	0.499	0.631	0.411	0.466	0.266
ambience-general (255)	0.799	0.819	0.738	0.571	0.531	0.645
drinks-style-options (32)	0.266	0.470	0.666	0.285	0.545	0.499
location-general (28)	0.285	0.444	0.444	0.000	0.000	0.000
drinks-quality (47)	0.521	0.444	0.399	0.285	0.545	0.235
restaurant-prices (80)	0.380	0.399	0.615	0.275	0.307	0.249
restaurant-miscellaneous (98)	0.344	0.230	0.153	0.133	0.226	0.099
service-general (449)	0.797	0.799	0.796	0.574	0.774	0.698
restaurant-general (442)	0.492	0.607	0.630	0.619	0.629	0.688
Micro F1 Average	0.495	0.516	0.535	0.407	0.449	0.458

PT-G=Pre-Trained-GloVe, CWL=Class Weighted Loss, DA=Data augmentation, PT-F=Pre-Trained-fastText

Experiments and Results - GermEval-2017 Data

- Our model performed better than the baselines and the GermEval best submission (BS).
- Our results are close to the current state of the art by Schmitt et al. (2018).

Variant	Macro F1					Micro F1
	average	*positive*	*negative*	*neutral*	*n/a*	
Pipeline LSTM-F	-	-	-	-	-	0.297
End-to-end LSTM-F	-	-	-	-	-	0.315
Pipeline CNN-F	-	-	-	-	-	0.295
End-to-end CNN-F	-	-	-	-	-	0.423
E2E LSTM-F	0.55	0.164	0.432	0.596	0.975	0.383
E2E LSTM-F + NS	0.53	0.113	0.290	0.583	0.945	0.266
E2E LSTM-F + STF	0.56	0.209	0.436	0.598	0.976	0.384
MC baseline	-	-	-	-	-	0.315
GermEval baseline	-	-	-	-	-	0.322
GermEval BS	-	-	-	-	-	0.354

E2E=End to End, NS=Negative Sampling, STF=Self Trained fastText Embeddings, LSTM-F=LSTM with fastText, MC=Majority Class, BS=Best Submission

Experiments and Results - Organic Food Data

- Our model performed poorly with fine-grained version of the data.
    - Ambiguous attributes. "*You are eating potentially toxic impurities along with the minerals.*"
    - Conflicting annotations.

Variant	Macro F1					Micro F1
	average	*positive*	*negative*	*neutral*	*n/a*	
PT-F + CWL + Fine	0.36	0.082	0.063	0.097	0.981	0.059
PT-F + CWL + Coarse	0.39	0.197	0.122	0.231	0.966	0.129

PT-F=Pre-trained fastText embedding, CWL=Class Weighted Loss

Experiments and Results - Organic Food Data

- Majority of the aspects had 10 or fewer data points and hence it was really difficult for the model to learn to classify them.

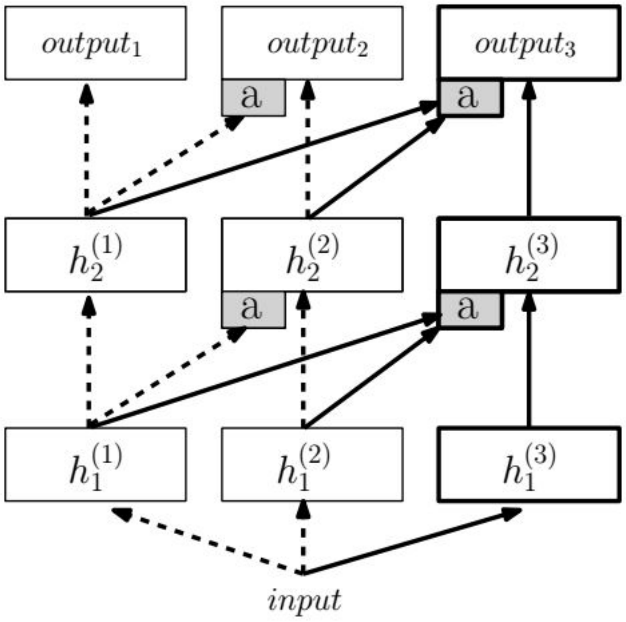
Count of data points (C)	Number of aspects with C data points	PT-F & CWL & Fine (Avg of per aspect micro F1 Score)
$0 < C \leq 10$	57	0.005
$10 < C \leq 20$	16	0.064
$20 < C \leq 50$	16	0.068
$50 < C \leq 100$	10	0.188
$C > 100$	12	0.278
Overall Micro F1 Score	111	0.059

CWL=Class Weighted Loss, PT-F=Pre-Trained-fastText

Transfer Learning

Transfer Learning - Progressive Neural Network (PNN)

- Progressive Neural Network (PNN) is a neural network architecture that improves upon some of the limitations of transfer learning - catastrophic forgetting.
- PNNs have lateral connections from a stack of previous networks trained on some source tasks.
- Weights of these lateral connections are learned by the current network. This gives the network some extra freedom to decide how much prior knowledge it wants to take from the previous tasks for the current task.



Source : Rusu et al. (2016)

Transfer Learning - Progressive Neural Network (PNN)

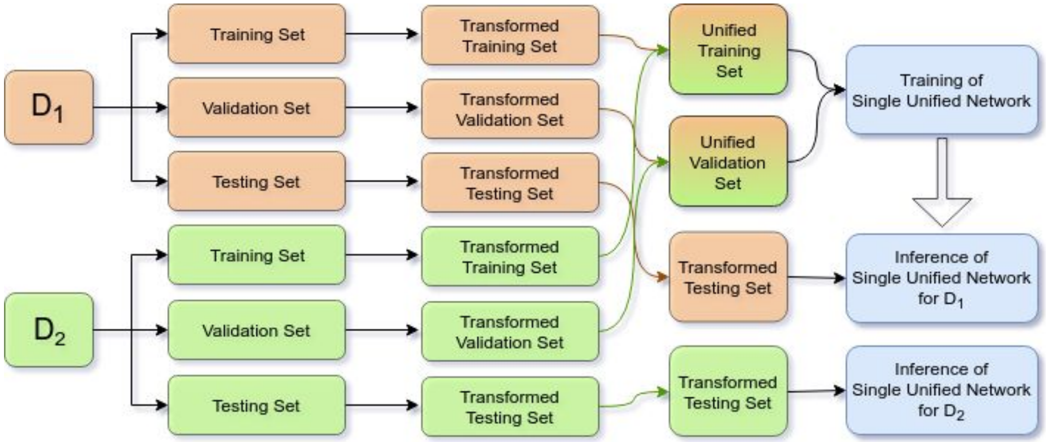
- PNN outperformed the baseline and fine-tuned version on restaurant data with an impressive 8 point margin.
- For laptops and organic data, we also saw an improvement of almost 3-4 points.

Variant	Restaurant PT-F CWL DA	Laptop PT-F CWL DA	Organic (Coarse) PT-F CWL
Baseline	0.394	0.246	0.129
Fine-tuned	0.324	0.229	0.062
PNN	0.471	0.276	0.170

PT-F=Pre-trained fastText embedding, CWL=Class Weighted Loss, DA=Data Augmentation, PNN=Progressive Neural Network

Transfer Learning - Single Unified Network (SUN)

- We merge the transformed data points from different domains to create a single unified dataset.
- Since our model can be trained on this single unified dataset so, we refer it as the Single Unified Network (SUN)



Transfer Learning - Single Unified Network (SUN)

- Two unified training datasets
  - restaurant+laptops+organic
  - restaurant+laptops
- With the first dataset our model performed poorly compared to baseline for restaurant & laptop domains. Performance was improved for organic domain.

Variant	Baseline	Fine-tuned	PNN	SUN (without organic)	SUN (with organic)
Restaurant PT-F CWL DA	0.394	0.324	0.471	0.399	0.353
Laptop PT-F CWL DA	0.246	0.229	0.276	0.260	0.224
Organic (Coarse) PT-F CWL	0.129	0.062	0.170	-	0.145

PT-F=Pre-trained fastText embedding, CWL=Class Weighted Loss, DA=Data Augmentation, PNN=Progressive Neural Network, SUN=Single Unified Network

Transfer Learning - Single Unified Network (SUN)

- With the second dataset our model performed better compared to baseline for restaurant & laptop domains.
- PNN outperformed Single Unified Network for all the three domains.

Variant	Baseline	Fine-tuned	PNN	SUN (without organic)	SUN (with organic)
Restaurant PT-F CWL DA	0.394	0.324	0.471	0.399	0.353
Laptop PT-F CWL DA	0.246	0.229	0.276	0.260	0.224
Organic (Coarse) PT-F CWL	0.129	0.062	0.170	-	0.145

PT-F=Pre-trained fastText embedding, CWL=Class Weighted Loss, DA=Data Augmentation, PNN=Progressive Neural Network, SUN=Single Unified Network

Future Work & Conclusion

Future Work

- Through investigation of results with the organic dataset.
- Many approaches in the literature claim to have achieved better results with CNNs than with RNNs or its variants. Going by the same spirit we feel it would be totally worth it to study how well a CNN variant of our approach performs.
- An interesting idea for generating a more effective aspect vector is by fusing various seed words that can define a given aspect.

Conclusion

- We proposed a novel architecture that is capable of jointly detecting aspects and corresponding sentiments.
- Another important contribution of this work was the input transformation technique. This technique provided some interesting benefits to the proposed model. It helped in addressing the following important problems.
    - Limit the number of output classes to 4.
    - By fixing the number of output classes, our model became capable of getting trained on datasets of different domains without requiring any architectural modifications. With this added capability we were able to transfer the knowledge from a source domain to a target domain using the approach of Progressive Neural Network (PNN) and Single Unified Network (SUN).

Conclusion

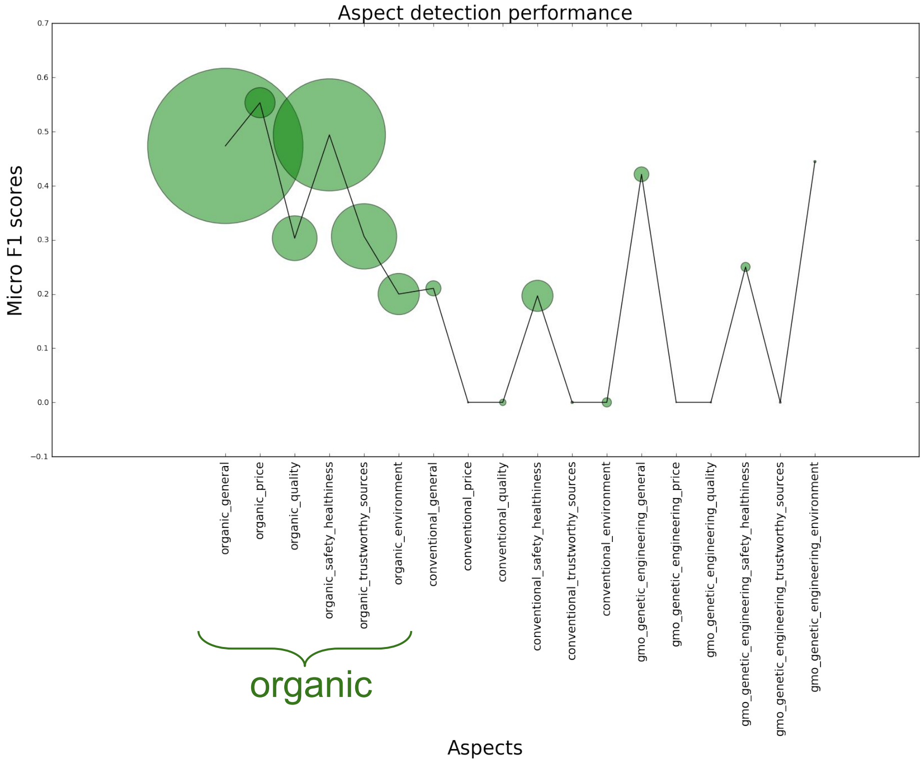
- To address the problem of limited and imbalanced data we implemented a data augmentation library (called simsent) that generates paraphrases of a sentence by replacing some of its words with synonyms or similar words.
- The results of our experiments with the GermEval-2017 dataset are very close to the current state of the art (Schmitt et al., 2018).
- To our best knowledge, we are the first ones to provide practical results on SemEval-2016 restaurant and laptop dataset for aspect-based sentiment analysis using a joint approach. These results can now act as a new baseline for future research in this area.

Thank you

Questions?

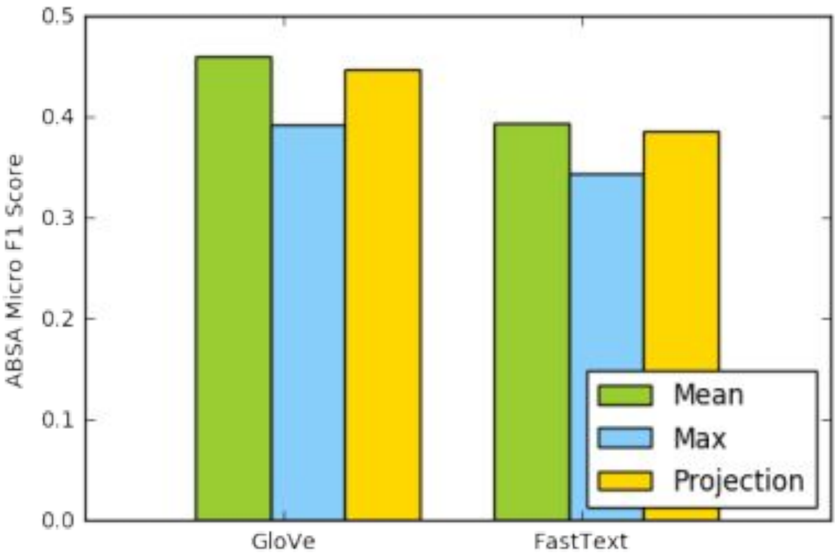
Experiments and Results - Organic Food Data

- Aspect detection scores for coarse-grained organic data. Overall micro f1 score was 0.237.
- Aspects with “*organic*” entity relatively had more data points and better accuracies compared to other aspects.



Experiments and Results - Aspect Fusion Techniques

- Mean and the projection fusion approaches performed better than max fusion for both GloVe and fastText.



Data Generation

	SemEval 2016 Restaurant	SemEval 2016 Laptop	GermEval 2017 Deutsche Bahn	Organic Fine grained	Organic Coarse grained
Review count	350	450	19432	962	962
Review count with IT	4200	52200	388640	106782	17316
Review count with IT & DA	5654	55136	129299*	-	-
Sentence count	2000	2500	-	6973	6973
Entities count	6	22	-	9	3
Attributes count	5	9	-	14	6
All Aspects count	30	198	20	126	18
Valid Aspects count	12	116	20	111	18

IT=Input Transformation, DA=Data Augmentation with similar words, *= Here, we performed negative sampling of the majority class

Table 5.1.: This table reports the statics of training set of various datasets.

Data Generation

	Positive	Negative	Neutral	N/A
Before Data Augmentation (Restaurant)	6.83%	3.08%	0.41%	89.66%
After Data Augmentation (Restaurant)	6.44%	2.91%	6.04%	84.58%
Before Data Augmentation (Laptops)	0.56%	0.37%	0.06%	98.99%
After Data Augmentation (Laptops)	0.55%	0.37%	1.06%	98.00%
Before Data Augmentation (DB)	0.54%	2.30%	2.64%	94.49%
After Data Augmentation* (DB)	1.61%	6.81%	7.82%	83.74%
Before Data Augmentation (Organic fine)	0.18%	0.16%	0.21%	99.43%
Before Data Augmentation (Organic coarse)	1.10%	1.01%	1.34%	96.53%

* = Here, we performed negative sampling of the majority class

Table 5.2.: This table shows the output class distribution of datasets for both before and after applying the data augmentation. Data augmentation was performed by generating similar sentences for SemEval-2016 Restaurant and Laptops dataset and for GermEval-2017 we did a negative sampling of majority class. Data augmentation was only performed for the neutral class. The first part of the table is for Restaurant data, the second part is for Laptop data, the third part for GermEval data, fourth part for Organic fine-grained data and finally the last part is for organic coarse-grained data.

Data Augmentation - Similar Sentences

Original Text	Augmentations using synonyms
For the price, you cannot eat this well in Manhattan.	For the monetary value, you cannot eat this well in Manhattan. For the Mary Leontyne Price, you cannot eat this well in Manhattan.
The food was lousy - too sweet or too salty and the portions tiny.	The nutrient was lousy - too sweet or too salty and the portions tiny. The food was lousy - too sweet or too salty and the fortune tiny. The food was lousy - too fresh or too salty and the portions tiny. The intellectual nourishment was lousy - too sweet or too salty and the portions tiny. The food was lousy - too sweet or too salty and the parcel tiny.
Ambiance- relaxed and stylish.	Ambiance- relaxed and stylish. Ambiance- relaxed and fashionable.
I'm very happy with this machine!	I'm very happy with this auto! I'm very happy with this automobile!
the features are great, the only thing it needs is better speakers.	the features are great, the only affair it needs is better speakers. the features are great, the only thing it needs is better verbaliser. the characteristic are great, the only thing it needs is better speakers. the features are great, the only thing it needs is well speakers. the features are great, the only matter it needs is better speakers. the feature of speech are great, the only thing it needs is better speakers. the features are great, the only thing it needs is better talker. the features are great, the only thing it needs is good speakers.
Great price and computer!	Great price and calculator! Great Mary Leontyne Price and computer! Great price and information processing system! Great monetary value and computer!

Table A.1.: Some examples of the augmentations that were generated by replacing words with synonyms.

Original Text	Augmentations using similar words
For the price, you cannot eat this well in Manhattan.	For the discount, you cannot eat this well in Manhattan. For the buy, you cannot eat this well in Manhattan.
The food was lousy - too sweet or too salty and the portions tiny.	The food was crappy - too sweet or too salty and the portions tiny. The food was shitty - too sweet or too salty and the portions tiny. The food was lousy - too sweet or too salty and the portions small. The food was lousy - too sweet or too salty and the portions large. The snacks was lousy - too sweet or too salty and the portions tiny. The cooking was lousy - too sweet or too salty and the portions tiny.
Ambiance- relaxed and stylish.	Ambiance- relaxed and fashionable. Ambiance- quiet and stylish. Ambiance- relaxed and sleek. Ambiance- easy-going and stylish. Ambiance- relaxed and trendy.
I'm very happy with this machine!	I'm very always with this machine! I'm very happy with this equipment! I'm very thankful with this machine! I'm very glad with this machine! I'm very happy with this Machine!
the features are great, the only thing it needs is better speakers.	the features are great, the only thing it needs is well speakers. the features are excellent, the only thing it needs is better speakers. the features are great, the only thing it needs is better subwoofers. the features are terrific, the only thing it needs is better speakers. the features are great, the only thought it needs is better speakers.
Great price and computer!	Great discount and computer! Amazing price and computer! Great pricing and computer! Wonderful price and computer! Great price and laptop!

Table A.2.: Some examples of the augmentations that were generated by replacing words with similar words.