Sentiment Analysis Using BERT and Multi-Instance Learning

Iryna Burak

Pablo Restrepo

July 6, 2020

Presentation Overview: What did we do on the last 2 weeks?

1. Generating BERT multilingual contextual embeddings for our sentences using the entire comment as context.
2. Updating and running our simple baselines.
3. Training out MilNet ignoring ‘neutral’ class.
4. Simple NN as baseline for our organic data.
5. Training our MilNet using the new contextual embeddings.
6. References.

1. Generating BERT multilingual contextual embeddings

Model used: BERT multilingual cased.

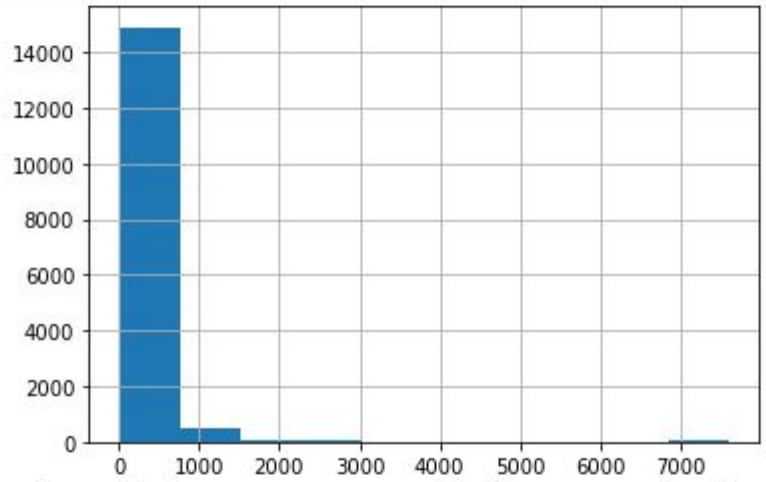
Method used to extract the embeddings:

- Generate contextual embeddings for each token, using the entire comment as context.
- For each sentence, use the mean of its tokens as embedding.

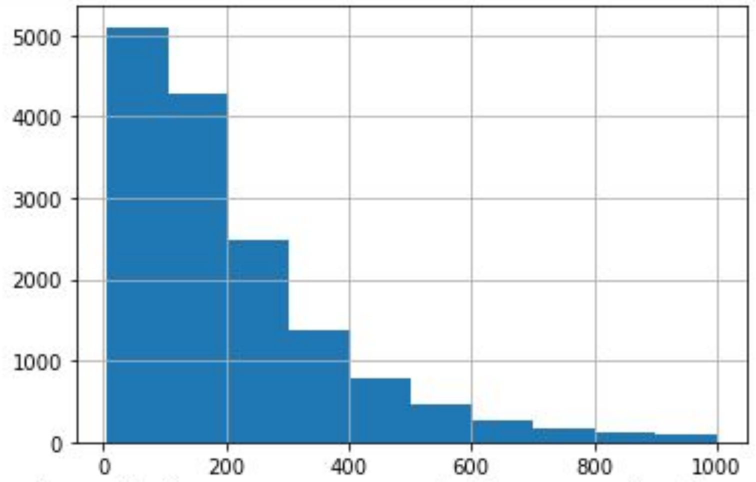
1. Generating BERT multilingual contextual embeddings

BERT can not handle sentences with more than 512 tokens, including [CLS] and [SEP].

For Amazon EN:



Number of tokens per comment before removing the outliers



Number of tokens per comment after removing the outliers

1. Generating BERT multilingual contextual embeddings

BERT can not handle sentences with more than 512 tokens, including [CLS] and [SEP].

For Amazon EN:

- Number of comments with more than **510** tokens: **1455**
- Total number of comments: **15561**
- Percentage of comments with more than 512 tokens: **9.3%**

- Total number of sentences: **131349**
- Number of sentences lost after removing comments with more than 510 tokens: **37817 (34%)**

1. Generating BERT multilingual contextual embeddings

BERT can not handle sentences with more than 512 tokens, including [CLS] and [SEP].

For Amazon DE:

- Only 2 comments had more than 510 tokens.

2. Updating and running our simple baselines

2.1 Baselines ignoring neutral class

Sentiwordnet

metric	amazon EN	organic
F1 (micro)	0.617	0.585
F1 (macro)	0.482	0.562

VADER

metric	amazon EN	organic
F1 (micro)	0.592	0.655
F1 (macro)	0.485	0.649

Textblob DE

metric	amazon DE
F1 (micro)	0.549
F1 (macro)	0.496

2. Updating and running our simple baselines

2.1 Baselines ignoring neutral class

NLTK sentiment analyzer

	F1 (micro)	F1 (macro)
amazon EN -> amazon EN	0.549	0.592
organic -> organic	0.482	0.578
amazon EN -> organic	0.607	0.394
amazon DE -> amazon DE	0.579	0.578

Scikit-learn SVM model

	F1 (micro)	F1 (macro)
amazon EN -> amazon EN	0.658	0.658
organic -> organic	0.640	0.622
amazon EN -> organic	0.614	0.473
amazon DE -> amazon DE	0.690	0.690

2. Updating and running our simple baselines

2.2 Baselines with neutral class

Sentiwordnet

metric	amazon EN	organic
F1 (micro)	0.329	0.384
F1 (macro)	0.273	0.376

VADER

metric	amazon EN	organic
F1 (micro)	0.473	0.480
F1 (macro)	0.340	0.442

Textblob DE

metric	amazon DE
F1 (micro)	0.371
F1 (macro)	0.366

2. Updating and running our simple baselines

2.2 Baselines with neutral class

NLTK sentiment analyzer

	F1 (micro)	F1 (macro)
amazon EN -> amazon EN	0.382	0.370
organic -> organic	0.45	0.437
amazon EN -> organic	0.356	0.187
amazon DE -> amazon DE	0.406	0.399

Scikit-learn SVM model

	F1 (micro)	F1 (macro)
amazon EN -> amazon EN	0.474	0.474
organic -> organic	0.456	0.436
amazon EN -> organic	0.378	0.282
amazon DE -> amazon DE	0.482	0.479

3. Training out MilNet ignoring ‘neutral’ class

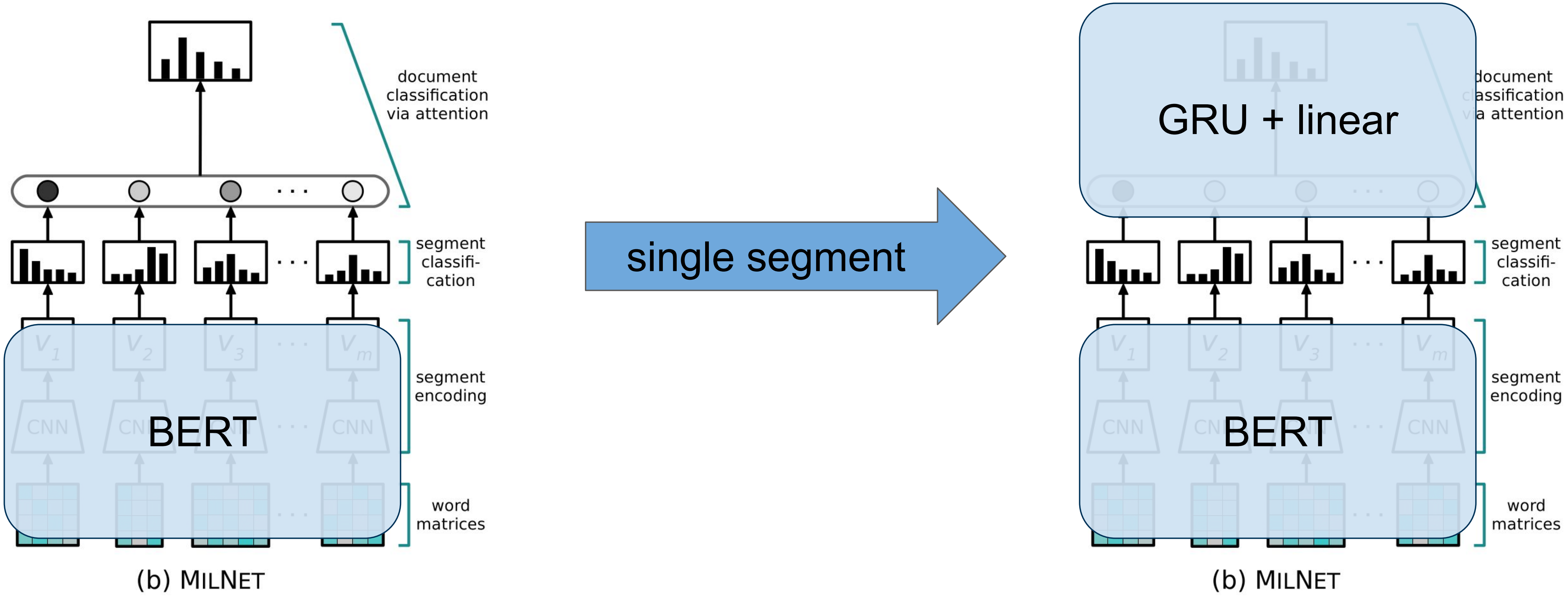
XLING without fine-tuning: 2 classes

metric	amazon EN	amazon DE	organic
F1 (micro)	0.813	0.820	0.643
F1 (macro)	0.804	0.820	0.640

XLING without fine-tuning: 3 classes

metric	amazon EN	amazon DE	organic
F1 (micro)	0.583	0.619	0.376
F1 (macro)	0.563	0.549	0.369

4. Simple NN as baseline for our organic data



Angelidis S. & Lapata M, Multiple Instance Learning Networks for Fine-Grained Sentiment Analysis, University of Edinburgh

4. Simple NN as baseline for our organic data

baseline: 2 classes

metric	organic
F1 (micro)	0.714
F1 (macro)	0.713

XLING with fine-tuning: 2 classes

metric	amazon EN	amazon DE	organic
F1 (micro)	0.725	0.775	0.704
F1 (macro)	0.705	0.774	0.703

baseline: 3 classes

metric	organic
F1 (micro)	0.612
F1 (macro)	0.612

XLING with fine-tuning: 3 classes

metric	amazon EN	amazon DE	organic
F1 (micro)	0.467	0.551	0.543
F1 (macro)	0.440	0.505	0.530

5. Training our MILNET using the new contextual embeddings

with fine-tuning
non-contextualized: 3 classes

metric	F1 (micro)	F1 (macro)
amazon EN	0.317	0.281
amazon DE	0.334	0.301
organic	0.539	0.519

with fine-tuning
contextualized: 2 classes

metric	F1 (micro)	F1 (macro)
amazon EN	0.630	0.629
amazon DE	0.584	0.499
nc organic	0.704	0.701

with fine-tuning
contextualized: 3 classes

metric	F1 (micro)	F1 (macro)
amazon EN	0.429	0.377
amazon DE	0.444	0.316
nc organic	0.555	0.543

5. Training our MILNET using the new contextual embeddings

without fine-tuning
non-contextualized: 3 classes

metric	F1 (micro)	F1 (macro)
amazon EN	0.573	0.556
amazon DE	0.503	0.415

without fine-tuning
contextualized: 2 classes

metric	F1 (micro)	F1 (macro)
amazon EN	0.793	0.776
amazon DE	0.688	0.684
nc amazon DE	0.676	0.673

without fine-tuning
contextualized: 3 classes

metric	F1 (micro)	F1 (macro)
amazon EN	0.564	0.545
amazon DE	0.519	0.482
nc amazon DE	0.516	0.494

6. References

- Angelidis S. & Lapata M, Multiple Instance Learning Networks for Fine-Grained Sentiment Analysis, Institute for Language, Cognition and Computation School of Informatics, University of Edinburgh.