Sentiment Analysis Using BERT and Multi-Instance Learning

Iryna Burak

Pablo Restrepo

June 22, 2020

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

1. Dataset hunting
2. Exploring our new Amazon reviews dataset in German
3. Generating embeddings for our new dataset using XLING, XLM and XLM-RoBERTa
4. Training the network
5. Plan for the next two weeks
6. References

1. Dataset Hunting

The German dataset we were using did not have many reviews in our relevant categories:

Health & Personal Care	37
Grocery	2
Beauty	1

We found a new dataset that contains many more reviews in our relevant categories:

Grocery	2737
Beauty	7162

2. Exploring our new Amazon reviews dataset in German

Amazon reviews dataset in German:

- 200.000 reviews;
- 31 categories.

Total number of reviews in our relevant categories (Beauty and Grocery): **9.899.**

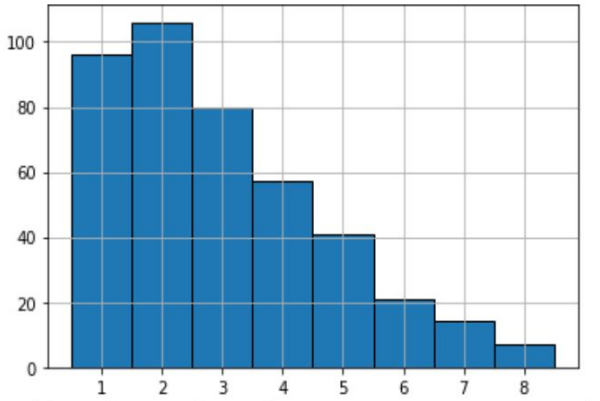
Number of sentences after splitting the comments (using NLTK sent tokenizer): **28.994.**

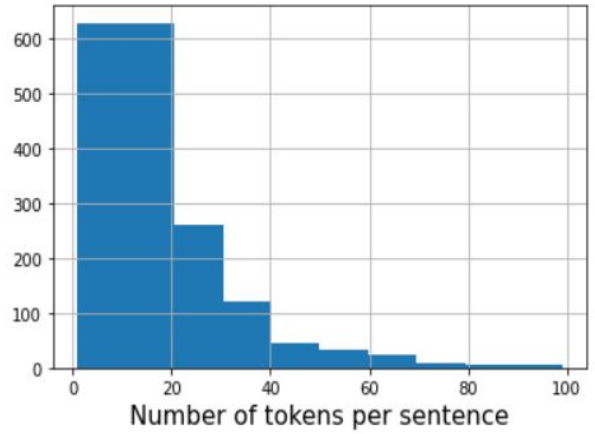
2. Exploring our new Amazon reviews dataset in German

The dataset is balanced:

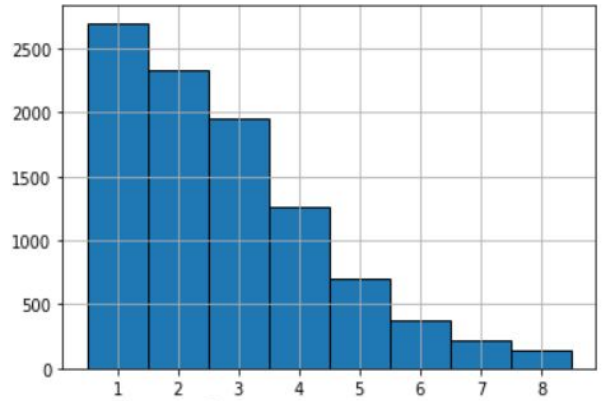
stars	number of instances
1	40000
2	40000
3	40000
4	40000
5	40000

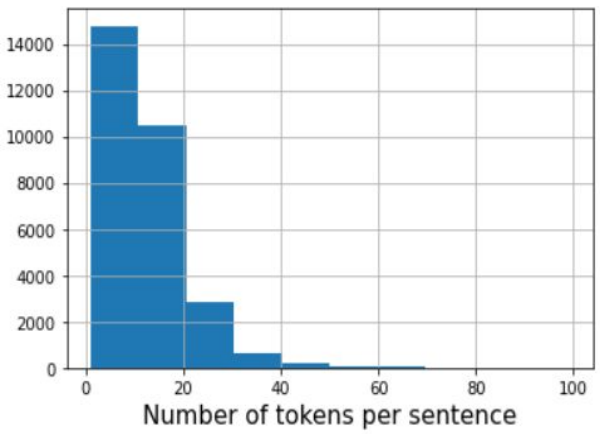
2. Exploring our new Amazon reviews dataset in German



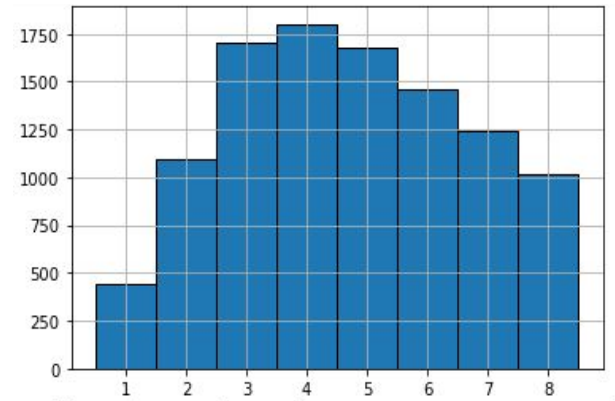


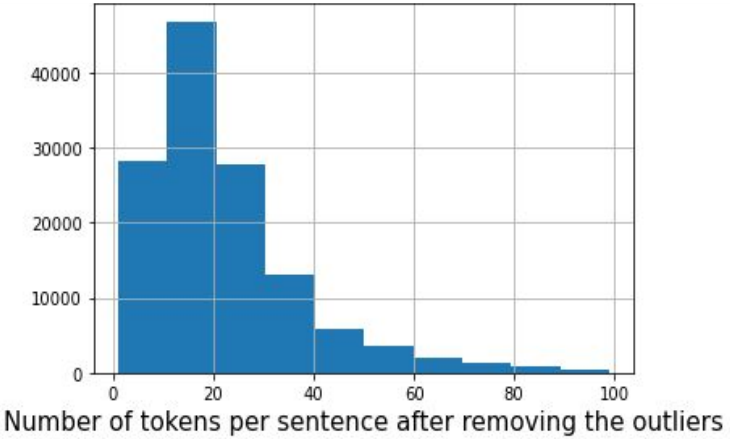
“Old” German Dataset





“New” German Dataset





English Dataset

3. Generating embeddings using XLING, XLM and XLM-RoBERTa

XLING:

Model: universal-sentence-encoder-xling-many (from tfhub)
Embeddings dimensionality: 512

XLM:

Model: xlm-mlm-ende-1024 (XLM Masked language modeling, English-German)(from Huggingface)
Pooling strategy: mean_tokens (not using CLS)
Embeddings dimensionality: 1024

XLM-RoBERTa:

Model: xlm-roberta-base (100 languages)(from Huggingface)
Pooling strategy: mean_tokens (not using CLS)
Embeddings dimensionality: 768

4. Training the network: base BERT

Amazon EN

True	-	0	+
-	71	14	54
0	37	10	99
+	21	5	162

Organic

True	-	0	+
-	87	38	21
0	49	118	32
+	19	47	92

Without fine-tuning on **organic**

metric	amazon EN	organic
F1 (micro)	0.674	0.498
F1 (macro)	0.664	0.494

With fine-tuning on **organic**

metric	amazon EN	organic
F1 (micro)	0.514	0.594
F1 (macro)	0.429	0.595

4. Training the network: multilingual BERT

Amazon EN

True	-	0	+
-	26	98	6
0	19	114	18
+	14	144	41

Amazon DE

True	-	0	+
-	2343	1758	41
0	1017	960	56
+	1367	2032	323

Organic

True	-	0	+
-	69	56	13
0	34	149	8
+	35	88	56

metric	amazon EN	amazon DE	organic
F1 (micro)	0.377	0.366	0.539
F1 (macro)	0.345	0.322	0.518

4. Training the network: XLING

Amazon EN

True	-	0	+
-	41	76	36
0	23	77	57
+	2	41	131

Amazon DE

True	-	0	+
-	1889	1639	614
0	593	769	671
+	274	870	2578

Organic

True	-	0	+
-	45	63	23
0	26	144	37
+	19	50	101

metric	amazon EN	amazon DE	organic
F1 (micro)	0.514	0.530	0.571
F1 (macro)	0.490	0.506	0.546

4. Training the network: XLM-RoBERTa

Amazon EN

True	-	0	+
-	11	117	8
0	5	109	23
+	4	128	74

Amazon DE

True	-	0	+
-	1270	2834	38
0	595	1352	86
+	328	2467	927

Organic

True	-	0	+
-	41	100	7
0	12	168	16
+	12	91	61

metric	amazon EN	amazon DE	organic
F1 (micro)	0.405	0.359	0.531
F1 (macro)	0.354	0.367	0.494

4. Training the network: XLM

Something is totally wrong!

Amazon EN
without fine-tuning on **organic**

True	-	0	+
-	0	0	111
0	0	2	160
+	0	3	204

Amazon EN
with fine-tuning on **organic**

True	-	0	+
-	0	111	0
0	0	162	0
+	0	207	0

5. Plan for the next two weeks

- Investigate the issue with XLM embeddings.
- Fine-tune our models.
- Improve the baselines.

6. References

1. M. Chidambaram, Y. Yang, D. Cer, S. Yuan, Y.-H. Sung, B. Strope, and R. Kurzweil. Learning Cross-Lingual Sentence Representations via a Multi-task Dual-Encoder Model. ArXiv e-prints, October 2018.
2. universal-sentence-encoder-xling-many, Google,https://tfhub.dev/google/universal-sentence-encoder-xling-many/1
3. arxiv2018-xling-sentence-embeddings, UKPLab, https://github.com/UKPLab/arxiv2018-xling-sentence-embeddings
4. Multi-lingual models, Huggingface, https://huggingface.co/transformers/multilingual.html