Sentiment Analysis Using BERT and Multi-Instance Learning

Iryna Burak

Pablo Restrepo

July 20, 2020

Presentation Overview:

1. Project Overview
2. Monolingual Sentiment Analysis
3. Cross Lingual Sentiment Analysis
4. Two-class Sentiment Analysis
5. Conclusions
6. References

1. Project Overview

Data, methods and experiments

1.1 Project Goal

Develop a domain specific sentiment analysis model to predict **sentence-level** sentiment on social media comments on organic food products.

1.1 Project Goal

Example:

I really love the product. It is really tasty and healthy. The only downside is that it is expensive. I would definitely buy it again.

- Overall sentiment of the comment: positive.
- But there are still some sentences with negative sentiment.

1.2 Data used

- **Amazon EN**, contains reviews in the categories:
  - Grocery and Gourmet Food
  - Health and Personal Care
  - Beauty

  We filtered the reviews that contained the word “organic”.

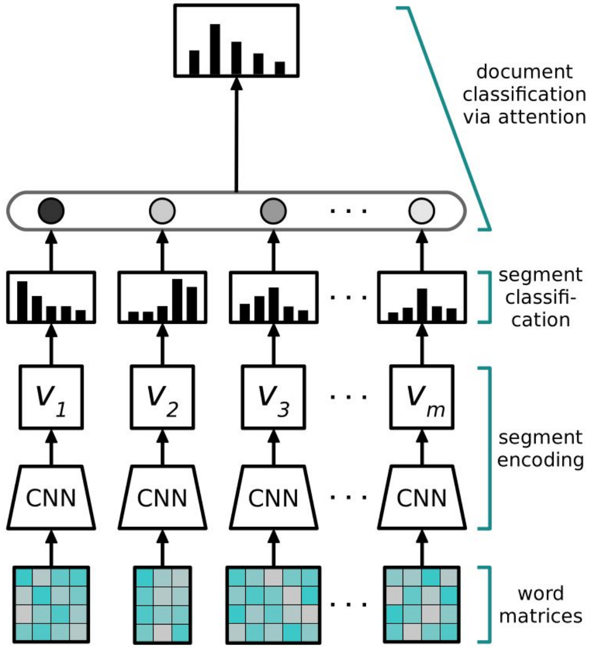
- Annotated **Organic** data:
  - Sentences about organic products annotated by domain experts.

- **Amazon DE**: contains reviews in the categories:
  - Beauty
  - Grocery

1.3 Tools and methods used

- Multi Instance Learning Networks:
  - Predict segment sentiments
  - Get a document sentiment via attention
  - Compute the loss with respect to **document**-level labels only

[1]



- Different embeddings as initial embeddings for our sentences and comments (BERT, RoBERTa, XLING).

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

1.4 Experiments

Train on	Fine-tune on	Test on
amazon EN	organic	organic
		amazon EN
organic	-	organic
amazon EN	-	amazon EN
		organic
		amazon DE

- Metrics: F1 scores (micro and macro)

1.5 Baselines

Baseline	English data	German data
Sentiwordnet	test	-
VADER	test	-
Textblob DE	-	test
NLTK Sentiment Analyzer (Naive Bayes)	train and test	-
Scikit-learn SVM model	train and test	-

2. Monolingual

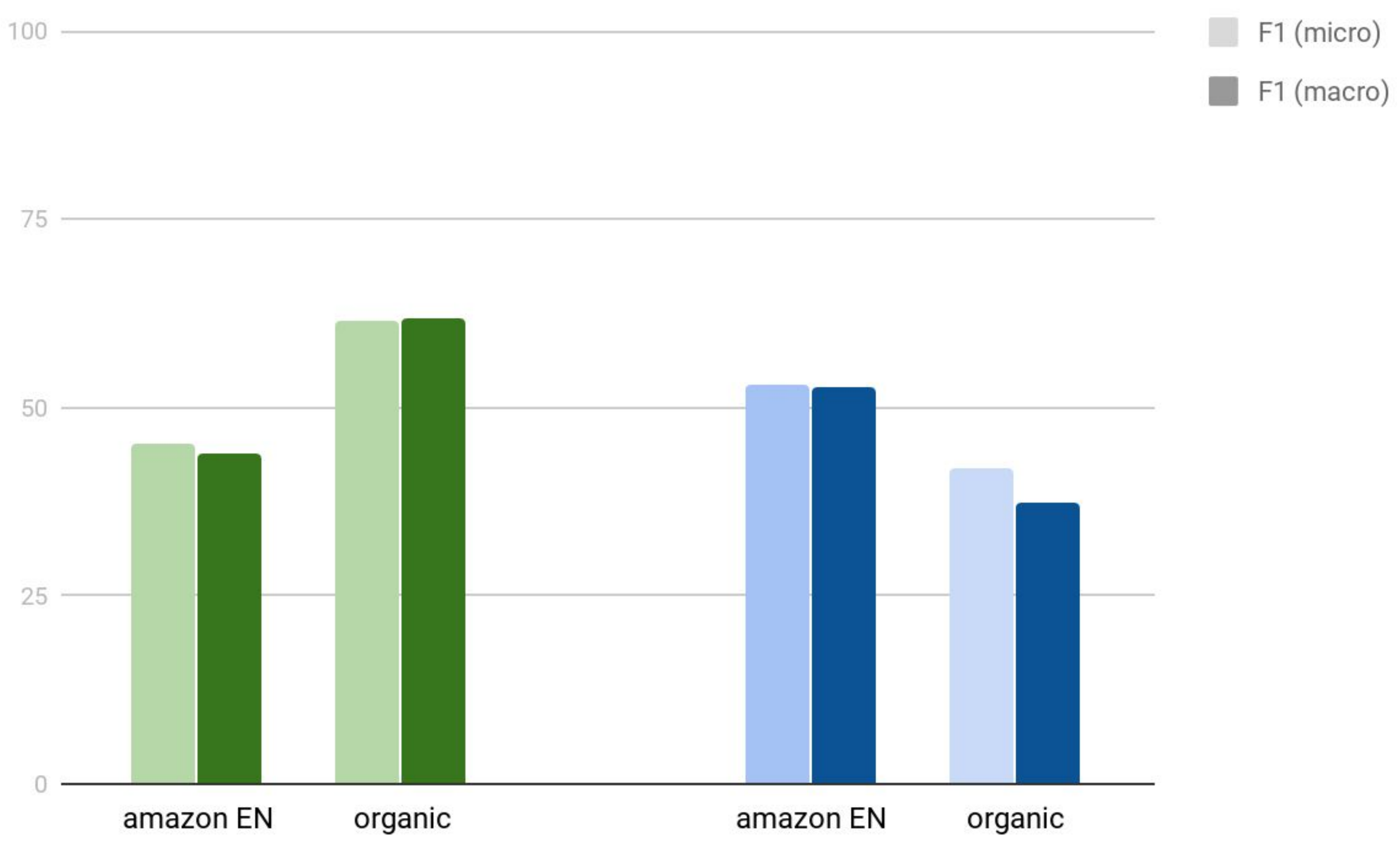
Plots and first results

2.1 Embeddings used

We used BERT base uncased (monolingual) to generate initial embeddings for our data.

Pooling layer: -2
Pooling strategy: reduce mean

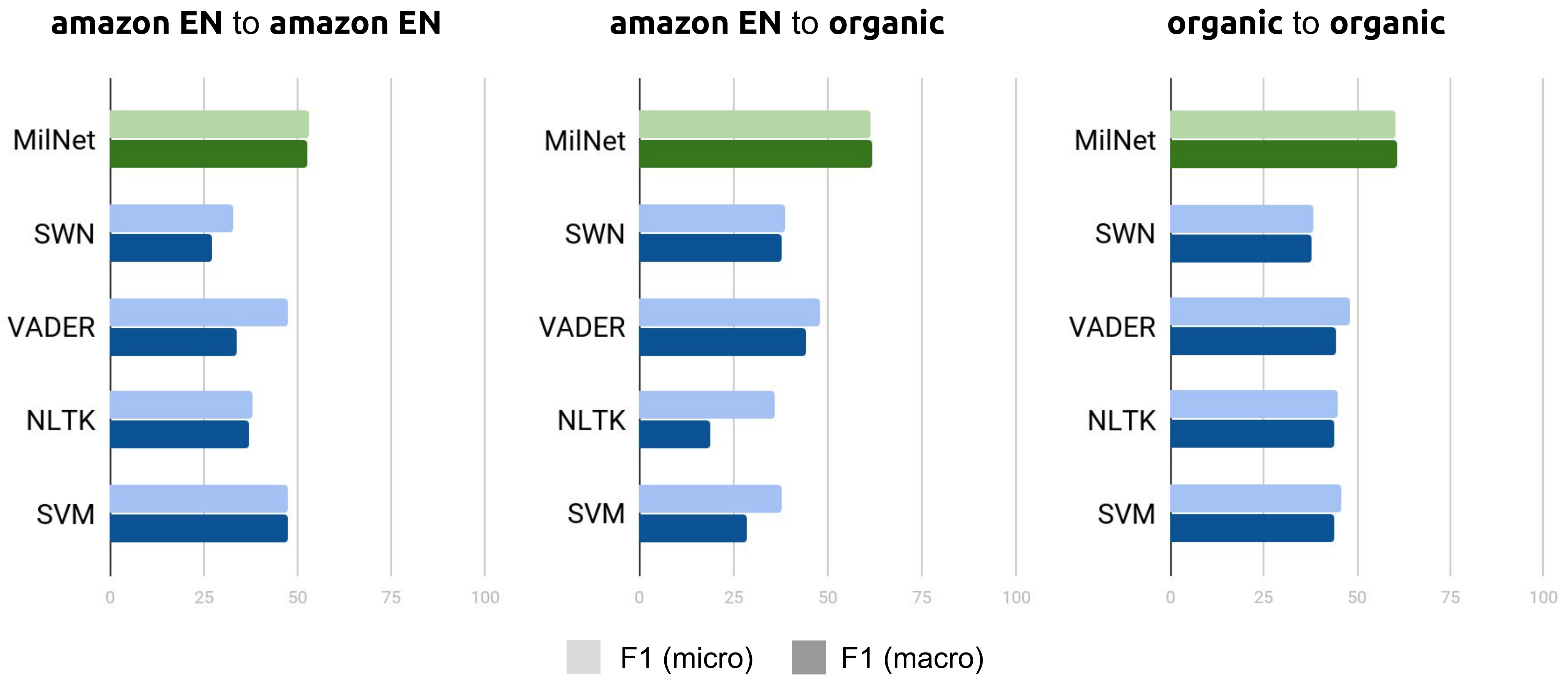
2.2 To fine-tune, or not to fine-tune



trained on: **amazon EN**, embeddings: **BERT base**

- Training on **amazon EN** only leads to poor results for **organic**.
- Fine-tuning on **organic** makes results for **amazon EN** significantly worse.
- **amazon EN** and **organic** contain different types of comments and/or annotated in different ways.

2.3 MilNet vs baselines



SWN = Sentiwordnet

3. Cross-lingual

New embeddings and new results

3.1 Embeddings used

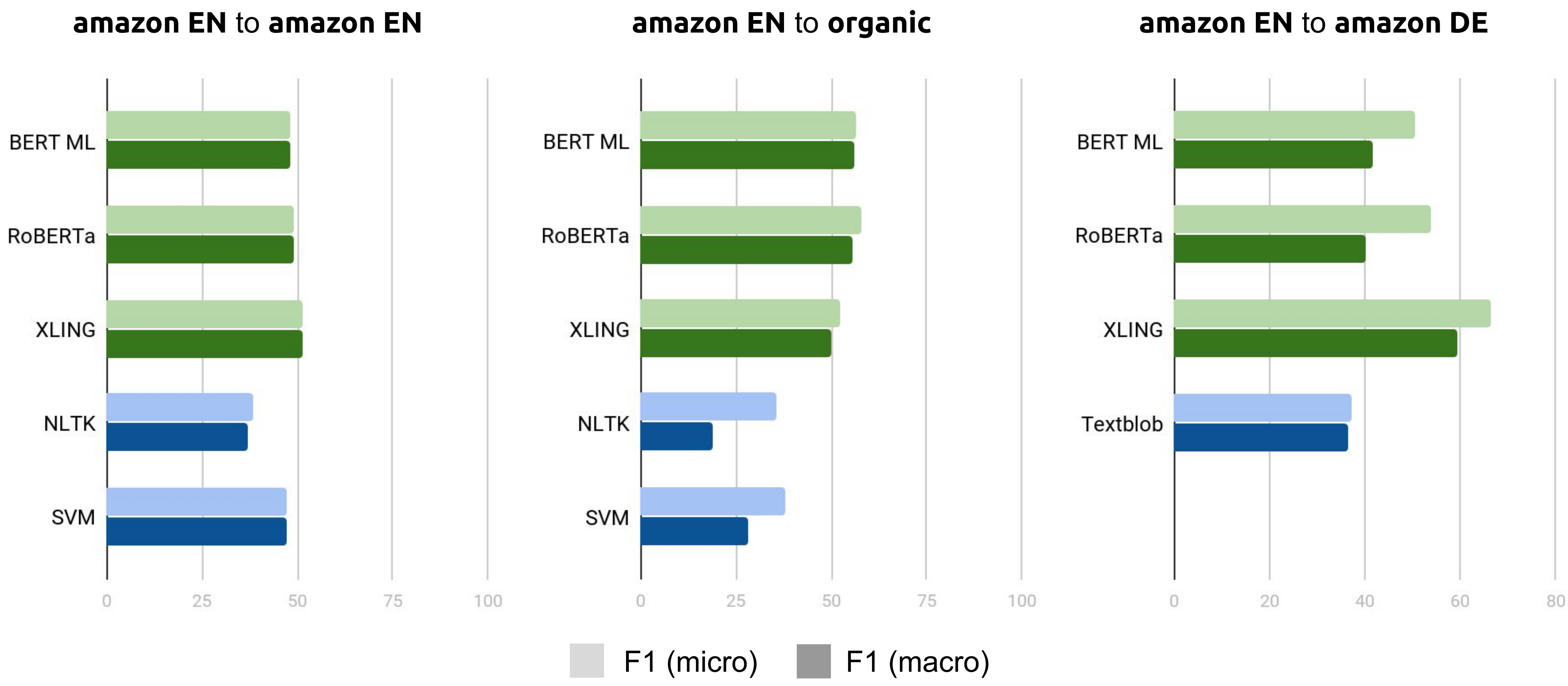
We used different multilingual embeddings:

- BERT base multilingual cased.
- RoBERTa
- XLING

For BERT and RoBERTa:

Pooling layer: -2
Pooling strategy: reduce mean

3.2 MilNet vs baselines



3.3 Sentence-level context vs comment-level context embeddings

We generated embeddings for each token of the comment using the entire comment as context.

Example:

[[CLS],'I', 'really', 'love', 'the', 'product', '.', 'It', 'is', 'really', 'tas', '##ty',[SEP]]

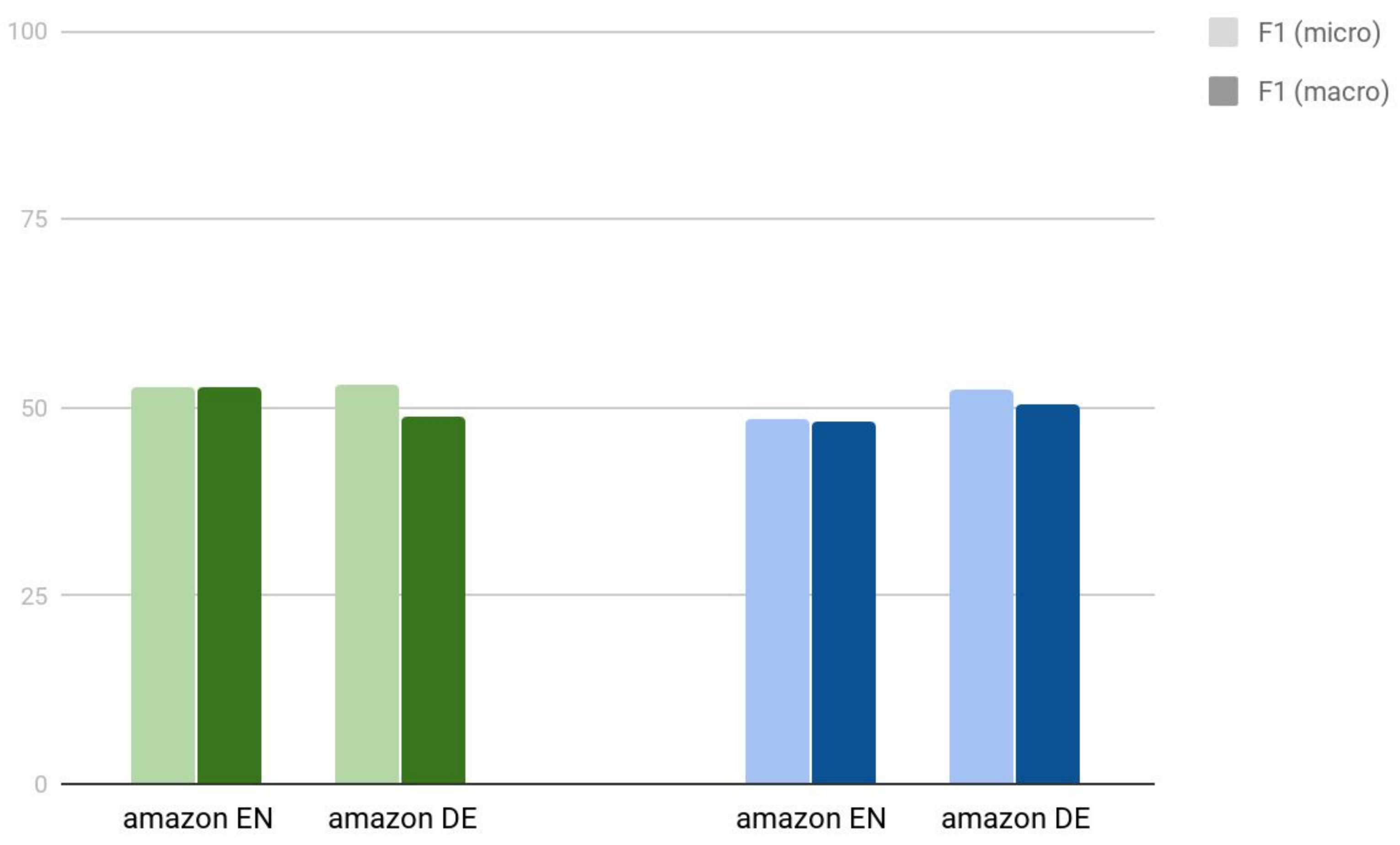
Sentence 1 embeddings: mean of the tokens: *'I', 'really', 'love', 'the', 'product',*
Sentence 2 embeddings: mean of the tokens: *'It', 'is', 'really', 'tas', '##ty'*

3.3 Sentence-level context vs comment-level context embeddings

BERT can only handle sentences with a maximum of **510** tokens + [CLS] and [SEP].

- Percentage of comments with more than 510 tokens: **9.3%**
- Number of sentences lost after removing comments with more than 510 tokens: **(34%)**

3.4 Comment-level vs sentence-level



trained on: **amazon EN**, embeddings: **BERT multilingual**

- To use comment-level context, we need to throw away large part of the data.
- But comment-level context helps a little bit with the classification.

4. Two-class analysis

More plots and even better results

3.1 Motivation

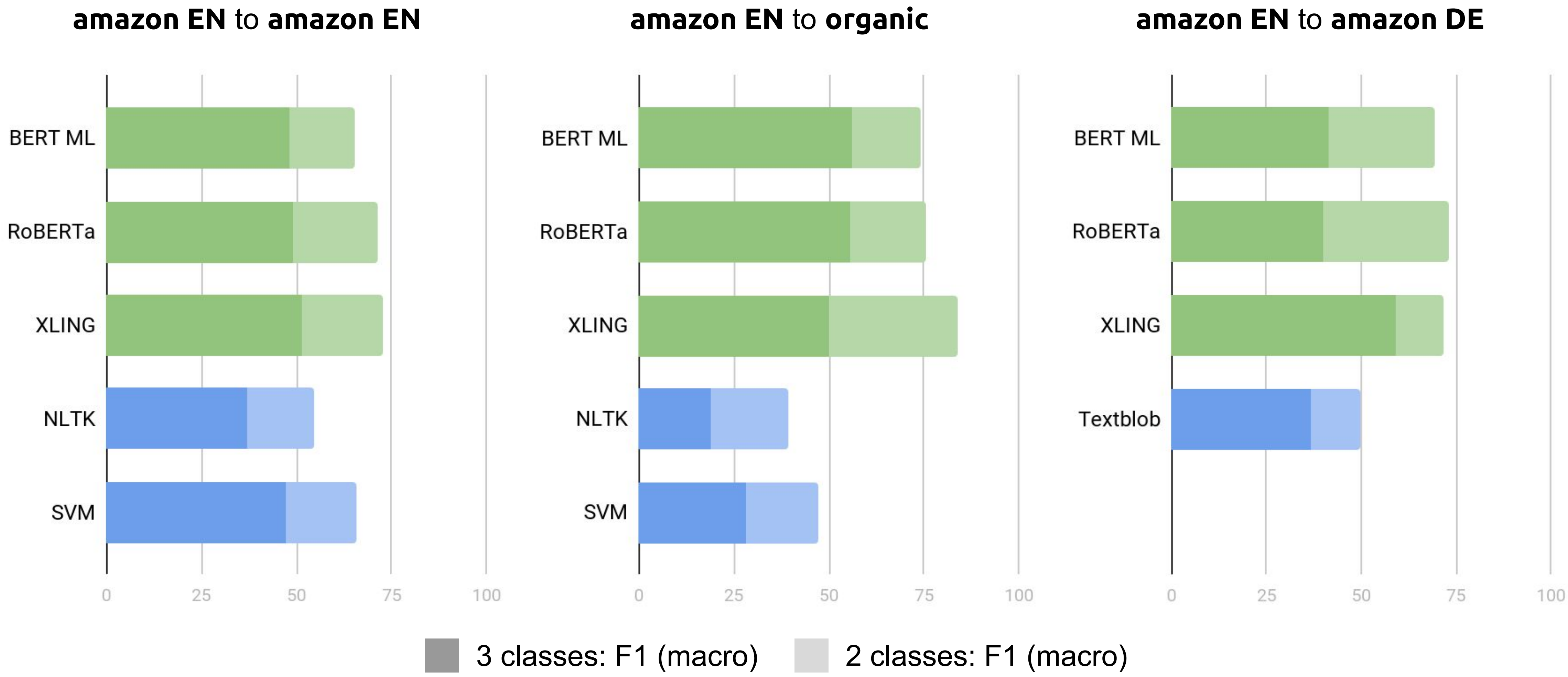
Confusion matrix

True	-	0	+
-	154	67	12
0	70	174	44
+	46	55	141

experiment: **amazon EN** to **organic**, embeddings: **BERT base**

- Neutral sentiment confuses the model.
- If the model is unsure about the decision, it predicts neutral sentiment.

3.2 MilNet vs baselines



5. Conclusions

The end

5.1 Conclusions

- MilNet works!
- The results for German data are better (possibly, Germans are just more explicit in their writing).
- In our setting, using the entire comment as context did not help a lot, but that result is not necessarily generalizable.
- In our setting, XLING was the best best choice for embeddings, but that result is not necessarily generalizable as well.

5.2 Conclusions

- Neutral sentiment was difficult for our models, but also for baselines!
    - Models choose neutral when unsure, but people also do this.
- Amazon data doesn’t use all the power of MilNet.
    - We don’t have sentence-level labels, only comment-level labels.
    - That is probably why **organic** gave better results than **amazon**.
- Amazon and organic datasets are different.
    - Even though datasets come from the same domain, they don’t help with analysis of each other.

6. References

[1] 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.

Thanks!