Comparing **Sentiment Analysis** of Discourse Units and Sentences
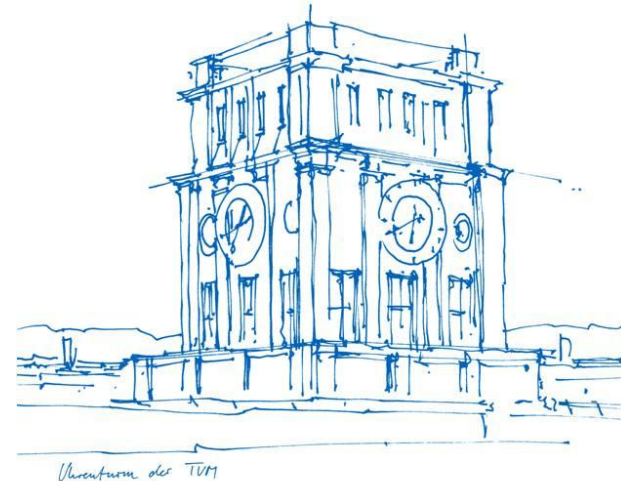
Group 9: Li Canchen, Hendrik Pauthner, Tim Pfeifle

Technische Universität München

Informatics

Research Group Social Computing

Munich, 17. July, 2019

DU vs Sentence

Review ➡ **Segments**

Sentence

```
1: The starters were quite bland.
2: Didn't enjoy most of them, but the
burger was brilliant!
```

```
The starters were quite bland.
Didn't enjoy most of them, but
the burger was brilliant!
```

Class: Positive

Rhetorical Structure Theory (RST) -Parser

Discourse-Unit (DU)

```
1: The starters were quite bland.
2: Didn't enjoy most of them,
3: but the burger was brilliant!
```
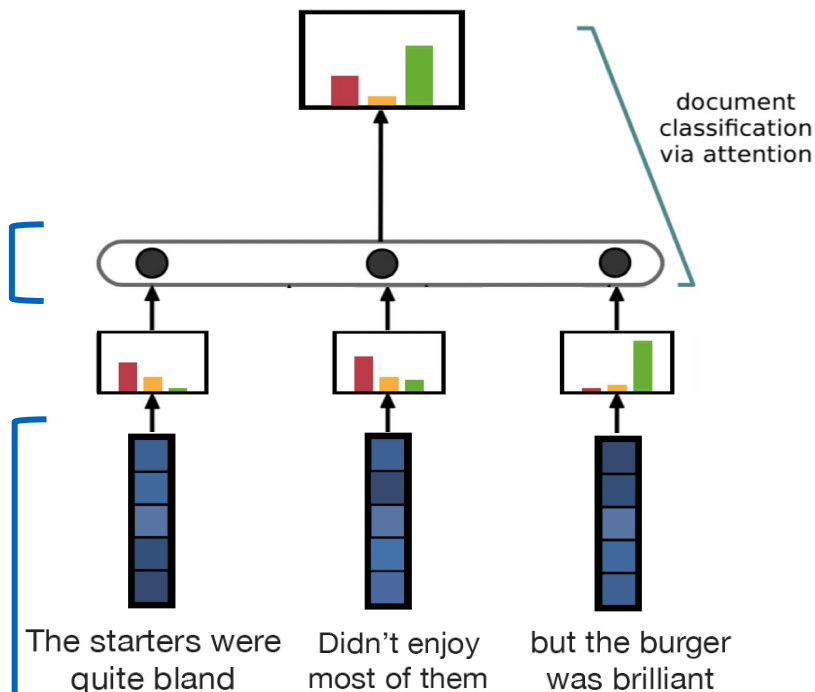
Multi-Instance Learning

Model

1) **Word** embeddings

2) **Segment** embeddings

document classification via attention

Attention

Segment Encoding

The starters were quite bland

Didn't enjoy most of them

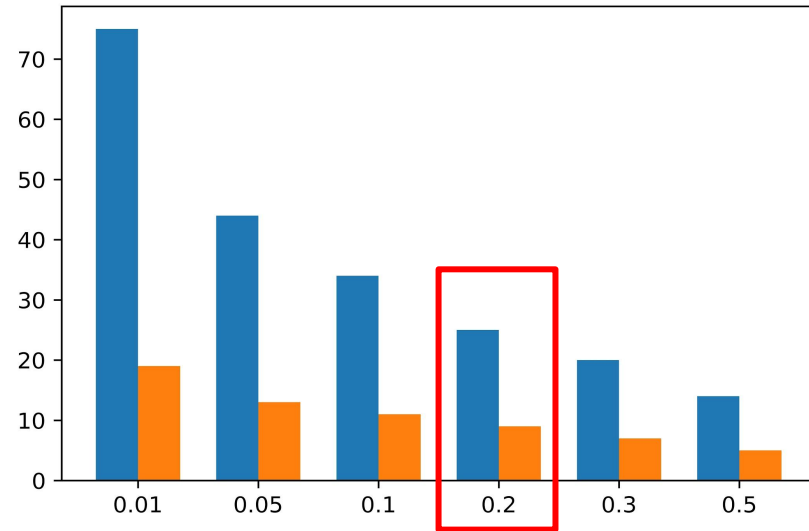but the burger was brilliant

4
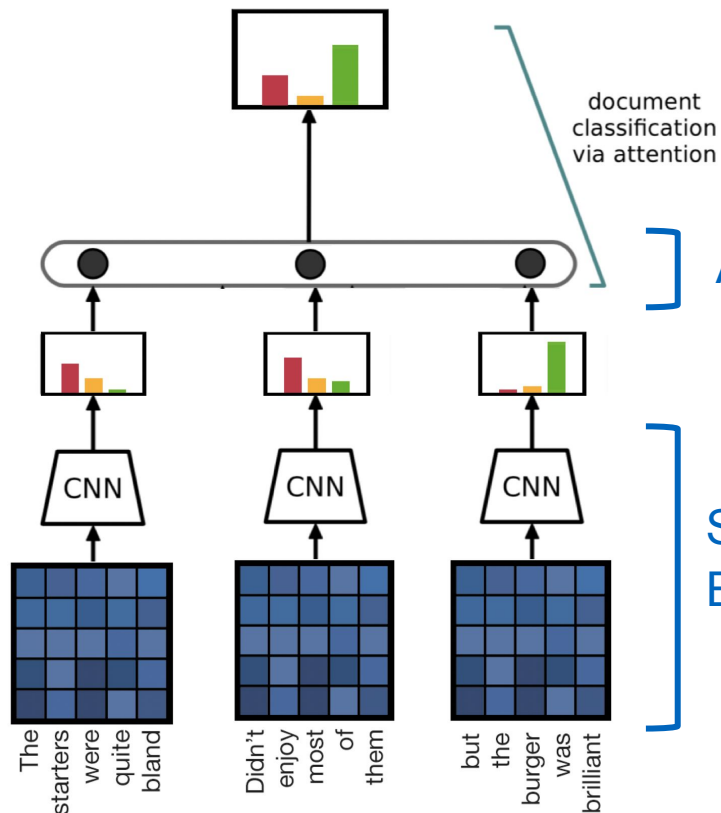
Datasets

Amazon Reviews: **Food**

- **Classes:** 3
    - Negative / neutral / positive
- **Distribution:** Balanced
- **Size:** 290 000 Reviews
- **Padding:**
    - Max-Seg: 10
    - Max-Words: 24
- **Scores:** Test set

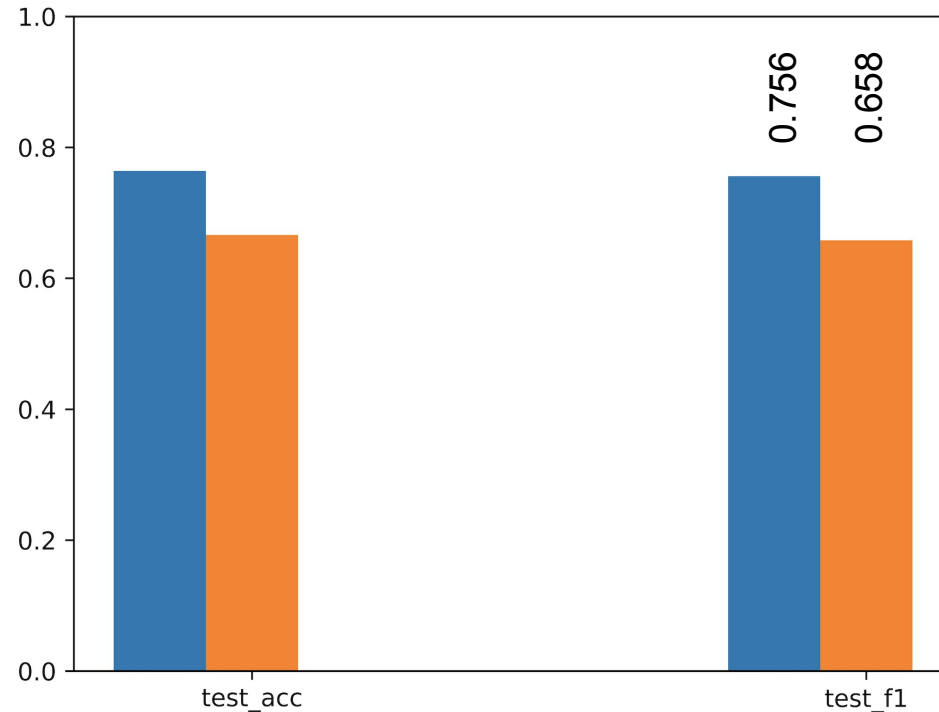💙 : # of Segments per Review

🧡 : # of Words per Review

Portion of exceeding reviews

Need fixed sizes → Zero-padding

Results: Different **Datasets**
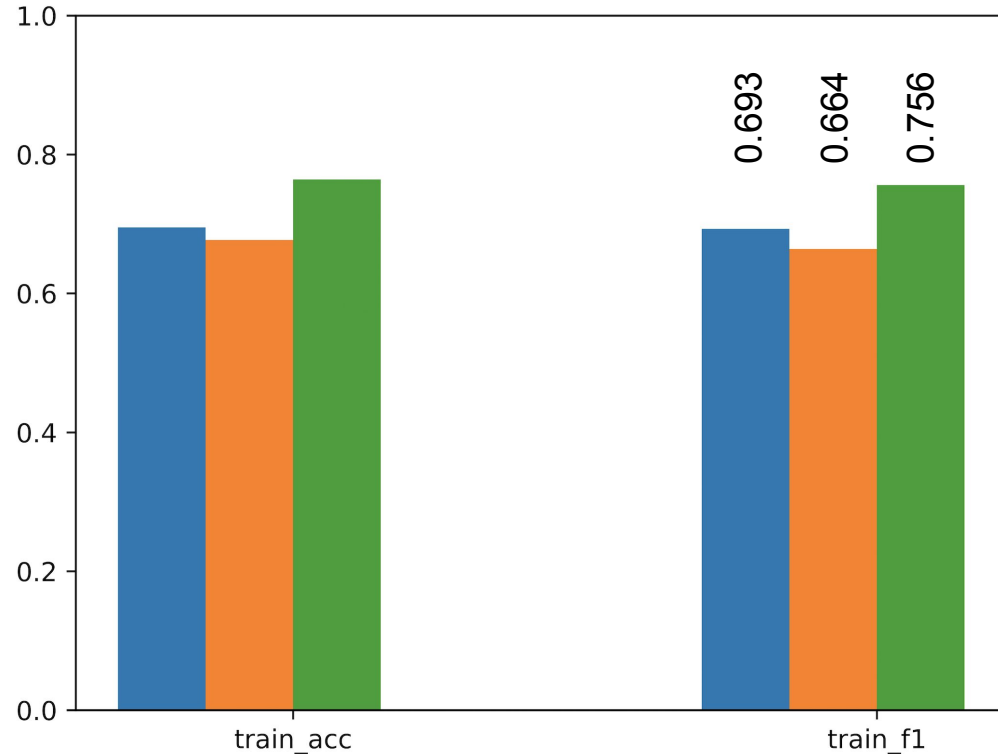
💙: **Food**

🧡: Electronics

Results: Different **Models**

💙: Naive Bayes

🧡: Without Attention

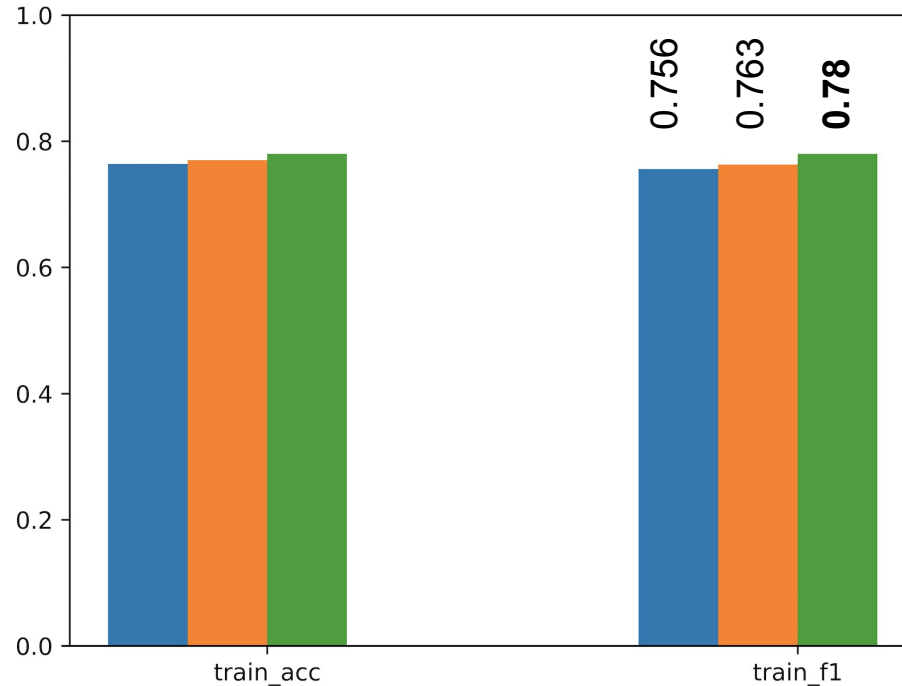💚: **With Attention**

Results: Different **Embeddings**

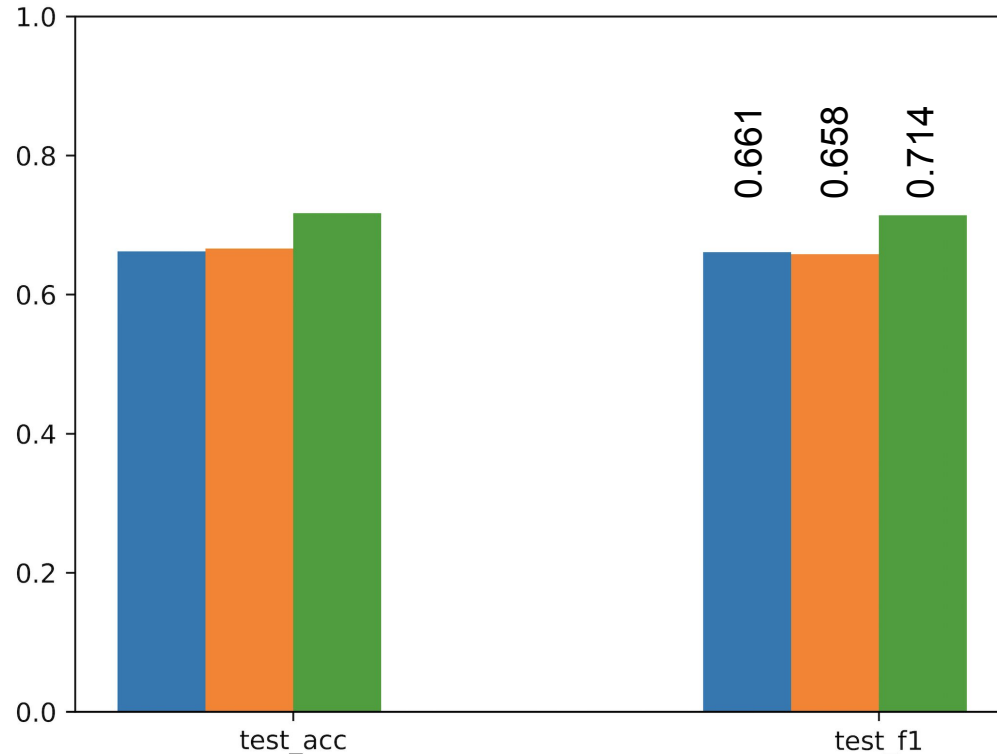💙: fastText

🧡: Google w2v

💚: **XLING**

Results: Different **Segment Levels**

💙: Document

🧡: Sentence

💚: **EDU**

On different dataset:
Electronics

Multi-Instance Learning

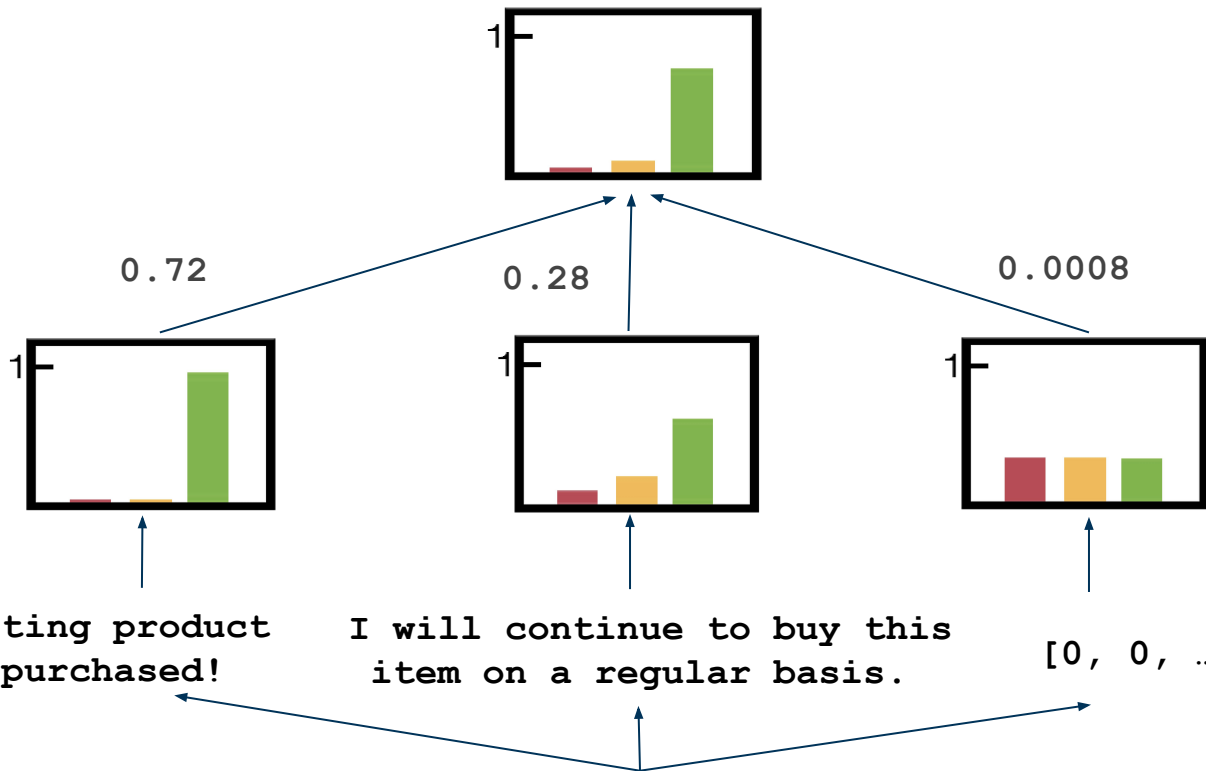Document-Label

Segment-Label

0.72   0.28   0.0008

The best tasting product
I have ever purchased!

I will continue to buy this
item on a regular basis.

[0, 0, …]   ...

The best tasting product I have ever purchased! I will
continue to buy this item on a regular basis.

Q&A

Goals for our Project

EDU Extraction + Preprocessing	Multi-Instance Sentiment Analysis	Comparison: Sentence-level vs EDU-level

Amazon Dataset EDU

Organic Dataset
→ Annotation Task

Multi-Instance Learning Model

Baseline (Naive Bayes)

Comparison
→ Sentence-wise vs EDU-wise features

On different Datasets
- Electronics
- Gourmet
- Organic Dataset