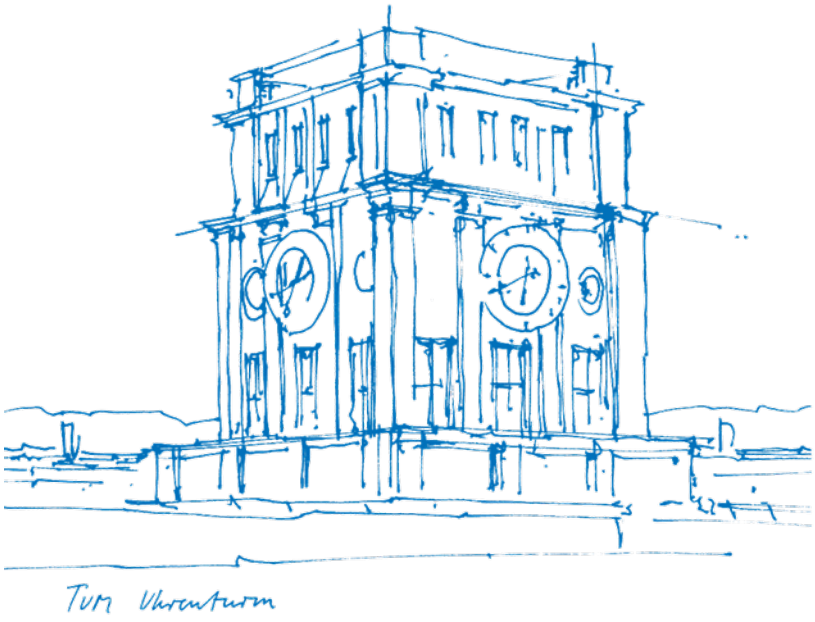
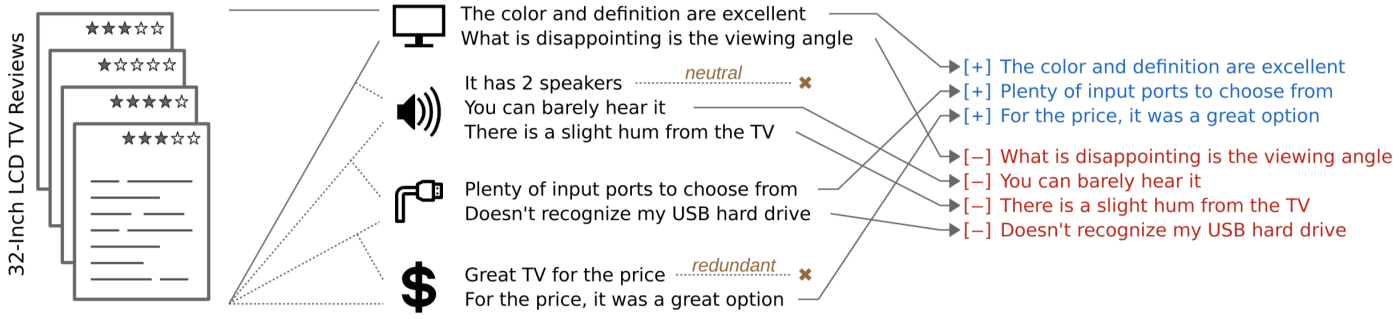
Comparing Aspect-Based Sentiment Analysis Classification of Discourse Units and Sentences

LI Canchen, Hendrik Pauthner, Tim Pfeifle
Technische Universität München
Informatics
Research Group Social Computing
Munich, 08. May 2019



Steps for Sentiment Analysis

1. Segment-Parsing
   a) Sentences
   b) Discourse units **(from RST parser)**
2. Aspect Extraction
   XLING as segment wise features
3. Sentiment Prediction
   Multi-Instance Learning (MILNET [1])
4. Summarization
   Identify aspect-specific and salient units + minimize redundancy



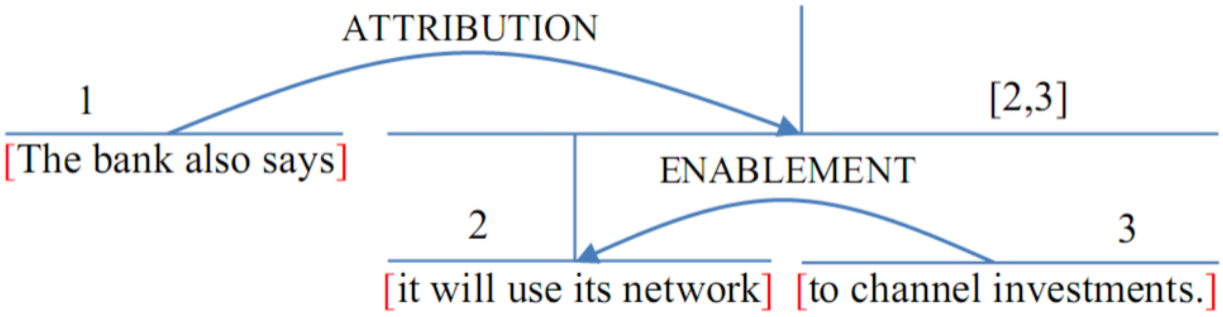
[1] Angelidis, Stefanos. "Multiple Instance Learning Networks for Fine-Grained Sentiment Analysis", 2018

RST Segmentation and EDUs

Goal: split text into elementary discourse units (EDUs) and analyze the rhetorical relationship between them.

Steps:

- **Segmentation** divide text into EDUs. Each EDU has a distinct purpose in its sentence
- **Parsing** Build rhetorical tree with the EDUs. It defines the relationships between the EDUs, thus helps to understand the text in a higher dimension



RST Segmentation Methods: Models

Rule-Based Model

Split the sentence into discourse units with fixed rules, including discourse cue, punctuation, grammar rules, etc.

Statistical Model

Consider the problem as a binary classification problem. For each word N_w, calculate the following probability of the word being a discourse unit boundary:

$$P(b|w,t) = \frac{Cnt\left(N_p \rightarrow ...N_w \uparrow N_r...\right)}{Cnt\left(N_p \rightarrow ...N_w N_r...\right)}$$

N_p is the parent of N_w in lexical syntactic tree, and N_r is the sibling of N_w. Split the sentence at the position of the words which have $p(b|w,t) > 0.5$.

RST Segmentation Methods: Models

Neuronal Network Model

Like the statistical model it approaches the problem as a binary classification and extracts similar features. Models the discourse segmentation with a neuronal network.

Sequential Model

The sequential model approaches the segmentation problem as a sequence labelling task conditioning the probability of a label sequence y on the observed sequence x.

XLING Embeddings

Universal Cross-lingual Sentence Encoder (based on M. Chidambaram et al., 2018)

- **Input:** Text of variable length in arbitrary language (does not have to be specified beforehand)
- **Output:** Embedding vector of length 512

Properties of XLING:

- Optimized for greater-than-word length text
- Trained on different tasks like Conversation Response Prediction or Natural Language Inference
  - Embeddings are useful out-of-the-box for several applications
  - Possibly we can use existing pre-trained XLING embeddings
- Cross-lingual $\rightarrow$ ability to analyze e.g. English and German reviews with the same model

Next Steps & Open Questions

- Input-Data preprocessing
- Comparing the different RST-Parsers
- Determine how close we want to follow the provided paper for the “sentiment analysis classification"[2].

[2] Angelidis, Stefanos. SSummarizing Opinions: Aspect Extraction Meets Sentiment Prediction", 2018