Structured Sentiment Analysis

Ayan Kumar Singh(2022122)
Rahul Jha(2022389)
Hardik Sachdeva(2022193)

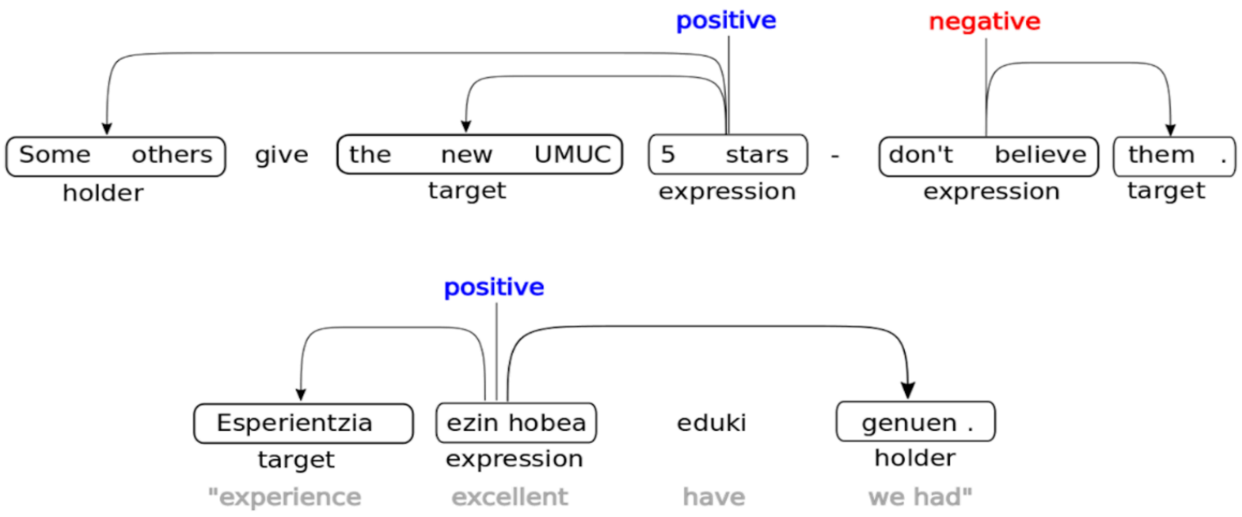
INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Problem Statement

Structured Sentiment Analysis

Structured Sentiment Analysis, where they attempt to predict the sentiment graphs, as seen above.

Formally, the task is to extract all of the opinion tuples O = Oi, ..., On in a text. Each opinion Oi is a tuple (h, t, e, p), where h is a holder who expresses a polarity p towards a target t through a sentiment expression e, implicitly defining the relationships between the elements of a sentiment graph.



Related Work

Structured Sentiment Analysis as Dependecy Graph parsing

- The paper explain about the structured sentiment problem as dependency graph parsing, where nodes are classifed and get hiearchial structures of sentiment holders, targets and expressions and where relation between them specifed by arcs.
  Paper : Link

Cross-Lingual and Multilingual Problems

- In multilingual Structured Sentiment Analysis (SSA), we have a big problem: not enough data for many low-resource languages. Models like *mBERT* and *XLM-R* can work across different languages, but they are not yet fully used for structured tasks like making sentiment graphs.
  Paper : Link

Datasets

Monolingual

- Norec (Norwegian professional reviews in multiple domains)
- Multibooked_ca(catalan hotel reviews)
- Opener_en (English hotel reviews)
- Opener_es (Spanish hotel reviews)
- darmstadt_unis (English online university reviews)
- MPQA
- Multibooked_es(Basque hotel reviews)

Baseline

Sequence Labelling

- Sequence Labeling: we try to make three BiLSTM models to get the extract holders, targets and expression and then trains a relation prediction model with the helps of max polling to create contextualized representation.
- We uses the text , first element (either holder or target), sentiment expression where we then concatenated and passed to a linear layer followed by sigmoid function.
- Training helps to predicts whether two elements have relation or not, which converts to classification problem.
- Combined all of these predictions, form a tupple of (h,t,e,p).

Metrics

Sequence Labelling

- For monolingual task, we have observed 0.559 sf1 score across average of all dataset, SP is 0.672 and SR is 0.65.
- For cross-lingual task, SF1 is 0.55, SP is 0.656 and SR is 0.63.

Table 3. Comprehensive Sequence-to-Sequence Labeling Results

Dataset	SF1	Precision	Recall
Darmstadt-unis	0.581	0.661	0.652
Multibooked-ca	0.530	0.650	0.620
Multibooked-eu	0.530	0.630	0.610
NoRec	0.541	0.684	0.651
Opener-en	0.610	0.700	0.690
Opener-es	0.600	0.700	0.690
MPQA	0.520	0.680	0.650

Table 4. Cross-Lingual Performance Evaluation

Dataset	**SF1**	**SP**	**SR**
Multibooked-ca	0.5300	0.6500	0.6200
Multibooked-eu	0.5300	0.6300	0.6100
Opener-es (Test)	0.6000	0.6900	0.6800

Datasets

Crosslingual

- Multibooked_ca(catalan hotel reviews)
- Opener_es (Spanish hotel reviews)
- Multibooked_eu(Basque hotel reviews)

Dataset	Language	# sents	# holders	# targets	# expr.
NoReC_fine	Norwegian	11437	1128	8923	11115
MultiBooked_eu	Basque	1521	296	1775	2328
MultiBooked_ca	Catalan	1678	235	2336	2756
OpeNER_es	Spanish	2057	255	3980	4388
OpeNER_en	English	2494	413	3850	4150
MPQA	English	10048	2279	2452	2814
Darmstadt_unis	English	2803	86	1119	1119

Methodology

Intial Processing and Tokenization

- Intially Structured Sentiment Analysis (SSA) extends basic sentiment analysis by extracting four key components from text to opinion is holders, Targets, expression, Polarity.
- We employ a pre-trained RoBERTa model (Liu et al., 2019) as our encoder, specifically “NbAiLab/nb-bert-base. The encoder transforms input text into contextual embedding
- We have compared the performance of RoBERTa with bert-base, bert-medium among them we have observed RoBERTa perform better.

Methodology

Tag Classification Layer

- A linear projection maps the contextual embeddings to tag logits:
- We use the BIO (Beginning-Inside-Outside) tagging scheme to represent the boundaries of opinion components, resulting in several possible tags: O, B-Source, I-Source, B-Target, I-Target, B-Polar_expression, I-Polar_expression.

$$T = WH + b \quad (1)$$

where $W \in \mathbb{R}^{d\times|Y|}$, $b \in \mathbb{R}^{|Y|}$, and $|Y|$ is the number of possible BIO tags. For a sequence of length n:

$$T \in \mathbb{R}^{n\times|Y|} \quad (2)$$

Methodology

Conditional Random Field

- After passing the data through Bert/ROBERTa, we have token-level hidden states, which are passed into a Conditional Random Field (CRF) layer, which helps to predict the BIO tags for quadruple (sources, targets, expression and polarity).
- CRF (Conditional Random Field) learns transition weights between tags, which helps to generate span detection while explicitly modelling constraints like ”I-Source cannot follow BTarget.”

$$\mathcal{L}_{\text{CRF}} = -\log\left(\frac{\exp(\text{score}(\text{gold\_tags}))}{\sum \exp(\text{score}(\text{all\_valid\_tags}))}\right)$$

Methodology

Polarity Classification

- A separate classification head with multiple layers determines the sentiment polarity (Positive, Negative, or Neutral) of identified opinions using the [CLS] token representation.
- Our model struggles to capture the mixed polarity sentences like the food was great, but the service was terrible.

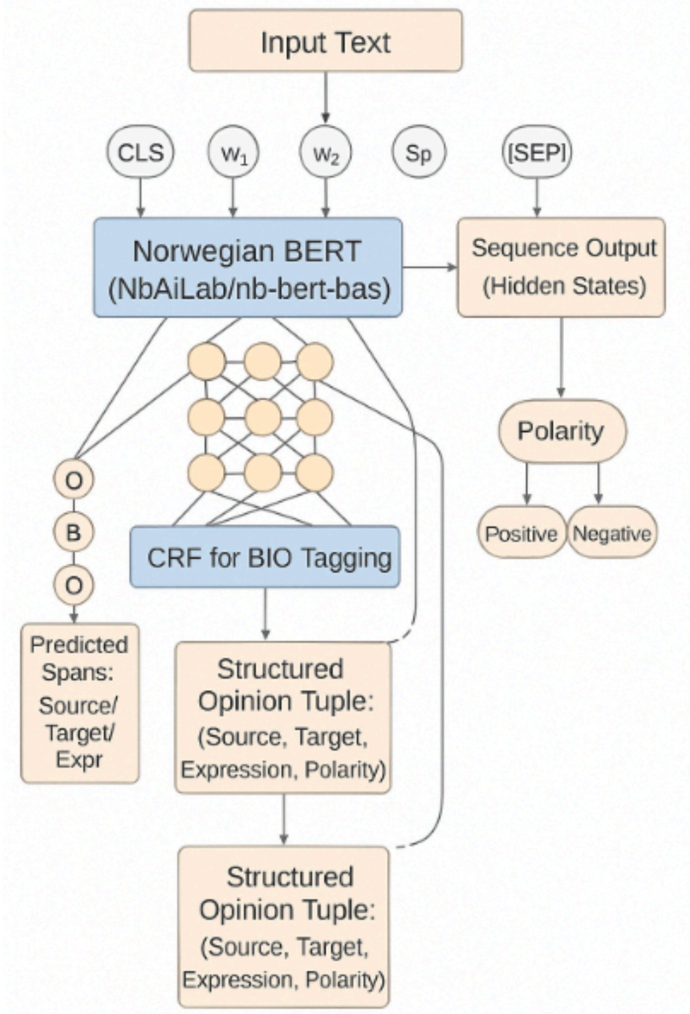


Figure 1. Model architecture for Structured Sentiment Analysis.

Methodology

Structured Opinion Generation

- In the final stage we have combined the CRF-predicted spans and the polarity classifier into the structured quadruplet form of target, holder, expression, and polarity as predictions of the problem-structured sentiment analysis.
- Span boundaries are converted to character offsets from tokens with the help of BERT’s offset mapping with adjacent merge.
- We also implemented the post-processing steps, filtering the spans which are shorter than 3 characters (which helps to remove noise and leads to the loss of a small portion of true predictions).

Result

Monolingual & Crosslingual

- For all datasets, we have observed an average of SF1 to be 0.46, SP to be 0.62 and SR to be 0.63, for cross-lingual SF1 is 0.55, SP is 0.65 and SR is 0.63.

Dataset	SF1	SP	SR
Opener-en	0.610	0.700	0.690
Opener-es	0.600	0.700	0.690
MPQA	0.520	0.680	0.650
NoRec	0.541	0.684	0.651
Multibooked-eu	0.530	0.630	0.610
Multibooked-ca	0.530	0.650	0.620

Dataset	**SF1**	**SP**	**SR**
Multibooked-ca	0.5300	0.6500	0.6200
Multibooked-eu	0.5300	0.6300	0.6100
Opener-es (Test)	0.6000	0.6900	0.6800

Result

Cross-lingual

- For all 3 datasets, we have observed average of SF1 to be 0.402 , SP to be 0.62 and SR to 0.62.

Table 2. Cross-Lingual Performance Evaluation

Dataset	**SF1**	**Precision**	**Recall**
Opener-es	0.4000	0.6000	0.6200
Multibooked-ca	0.4001	0.6490	0.6310
Multibooked-eu	0.4071	0.6640	0.6310

Analysis

- Role of Bert, It improves the span extraction result which performs by CRF, It performs better result specially in English and similar language that has been found.
- We have observed that polarity predictions outperformed in Multibooked_eu, Multibooked_ca, it’s because we have observed that because of the number of domains and charecterstics of the data, while in other data contains complex expression and ambigous expression (different polarity on different context).
- We have seen better results in token level Bert representation as compare character level.

Further

- In the future, we would likely to go with new approach dependecy parsing and sentiment graph parsing, by augmenting the token-level representation with contextualized vector from thier heads in a depenecy tree (Kurtz et al., 2020).
- Exploring more on multitask-learning to get deeper understanding of deepency parse.
- Different graph parsing approaches eg Point to point Network, PERIN((Samuel and Straka, 2020).
- In sequence span extraction, we can serialized the model using large pretrainned model to predict these serialized tupples.

THANK YOU!