UiO : Department of Mathematics
University of Oslo

FgFlex

A flexible, multitasking sequence-labeler for fine-grained sentiment analysis

Per M.C. Halvorsen

June 15, 2022

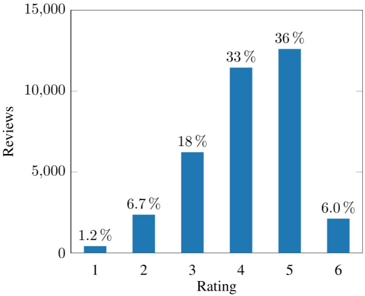
Table of contents

Table of contents

Understanding sentiment analysis

Annotation level

- Document level



Definition

Document level annotations

- a single score for a *large body of text*
- any number of sentences
- polarity score for the entire document


SESONG 1

Abbott Elementary

En engasjert og morsom arbeidsplasskomedie.

Publisert 09.06.22, kl 05.00.

Skrevet av
Sigurd Vik
sigurd.vik@nrk.no

OM SERIEN

PÅ DISNEY+: «Abbott Elementary» er en engasjert, lun og morsom arbeidsplasskomedie satt til en underfinansiert offentlig barneskole i Philadelphia, USA.

Serieskaper og hovedrolleinnehaver Quinta Brunson leder an i en lærergjeng det er umiddelbart lett å like, og som tar oss med inn i dagligdagse skoleutfordringer det er lett å sette seg inn i. Og selv om sesong 1 låner mye fra gode sjangerkollegaer, og

Understanding sentiment analysis

Annotation level

- Document level
- Sentence level

Definition

Sentence level annotations

- single score *per sentence*
- polarity, evaluative, or others

Dette	*gir*	*et*	*gjennomsnitt*	*på*	*27,3*
this	gives	an	average	on	27,3
MB/sek	*som*	*er*	*meget*	*bra.*	
MB/sec	which	is	very	good	

'This gives us an average of 27,3 MB / sec, which is very good.'

Det	*hele*	*var*	*også*	*lekkert*	*presentert.*
The	whole	was	also	tastefully	presented.

'Everything was tastefully presented.'

Understanding sentiment analysis

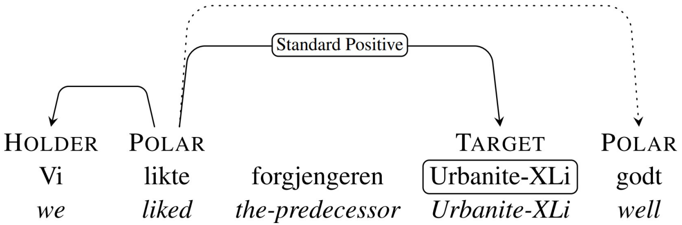
Annotation level

- Document level
- Sentence level
- Fine-grained level

Definition

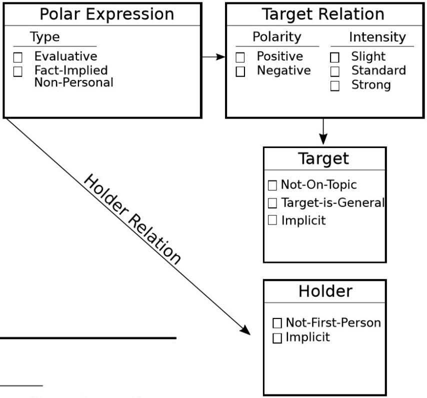
Fine-grained annotations

- label each opinion
- opinions include expression, target, holder, & polarity



NoReC$_{fine}$ dataset

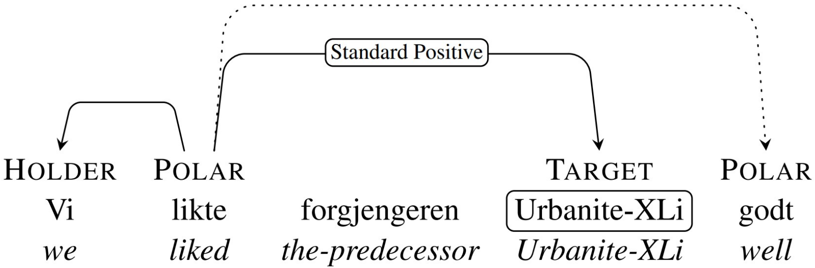
- By L.T.G. at UiO
- Part of S.A.N.T project
- Fundamental for this thesis



	# Examples				
	Train	Dev.	Test	Total	Avg. len.
Sents.	5915	1151	895	7961	16.8
Holders	584	76	75	735	1.1
Targets	4458	832	709	5999	2.0
Polar exp.	5659	1050	872	7581	4.6

Approaches to solve FGSA

- Structured sentiment analysis



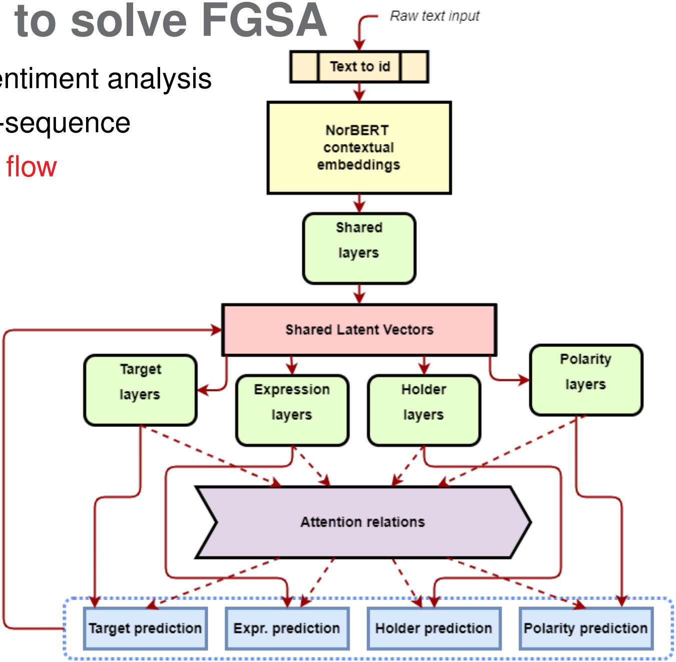
Approaches to solve FGSA

- Structured sentiment analysis
- Sequence-to-sequence

Text	Vi	likte	forgjengeren	Urbanite	XLi	godt
Expression	O	B	O	O	O	B
Holder	B	O	O	O	O	O
Polarity	O	O	O	+	+	O
Target	O	O	O	B	I	O

Approaches to solve FGSA

- Structured sentiment analysis
- Sequence-to-sequence
- General data flow



Architectures

Baselines

- IMN

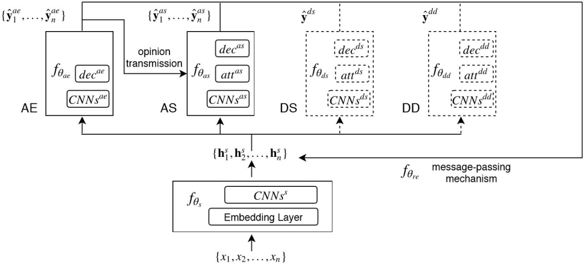
An Interactive Multi-Task Learning Network for End-to-End Aspect-Based Sentiment Analysis

Ruidan He[†‡], Wee Sun Lee[†], Hwee Tou Ng[†], and Daniel Dahlmeier[‡]
[†]Department of Computer Science, National University of Singapore
[‡]SAP Innovation Center Singapore
[†]{ruidanhe,leews,nght}@comp.nus.edu.sg
[‡]d.dahlmeier@sap.com

Abstract

Aspect-based sentiment analysis produces a list of aspect terms and their corresponding sentiments for a natural language sentence. This task is usually done in a pipeline manner, with aspect term extraction performed first, followed by sentiment predictions toward the extracted aspect terms. While easier to develop, such an approach does not fully exploit joint information from the two subtasks and does not use all available sources of training information that might be helpful, such as document-level labeled sentiment corpus. In this paper, we propose an interactive

treated separately and the overall task is performed in a pipeline manner, which may not fully exploit the joint information between the two tasks. Recently, two studies (Wang et al., 2018; Li et al., 2019) have shown that integrated models can achieve comparable results to pipeline methods. Both works formulate the problem as a single sequence labeling task with a unified tagging scheme[1]. However, in their methods, the two tasks are only linked through unified tags, while the correlation between them is not explicitly modeled. Furthermore, the methods only learn from aspect-level instances, the size of which is usu-



AE: *aspect term and opinion term co-extraction*

AS: *aspect-level sentiment classification*

DS: *document-level sentiment classification*

DD: *document-level domain classification*

Architectures

Baselines

- IMN
- RACL

Relation-Aware Collaborative Learning for Unified Aspect-Based Sentiment Analysis

Zhuang Chen, Tieyun Qian*
School of Computer Science, Wuhan University, China
{zhchen18, qty}@whu.edu.cn

Abstract

Aspect-based sentiment analysis (ABSA) involves three subtasks, i.e., aspect term extraction, opinion term extraction, and aspect-level sentiment classification. Most existing studies focused on one of these subtasks only. Several recent researches made successful attempts to solve the complete ABSA problem with a unified framework. However, the interactive relations among three subtasks are still under-exploited. We argue that such relations encode collaborative signals between different subtasks. For example, when the opinion term is *"delicious"*, the aspect term must be *"food"* rather than *"place"*. In order to fully exploit these relations, we propose a Relation-

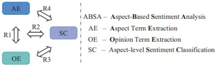
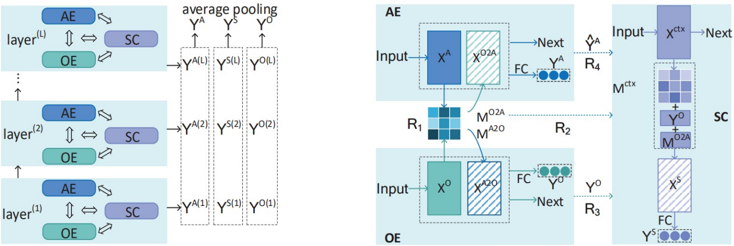


Figure 1: Interactive relations among subtasks in ABSA (left), and a list of abbreviations (right).

Most existing works treat ABSA as a two-step task containing AE and SC. They develop one separate method for each subtask (Tang et al., 2016; Xu et al., 2018; Li et al., 2018a; Hu et al., 2019), or take OE as an auxiliary task of AE (Wang et al., 2017; Li et al., 2018b). In order to perform ABSA for practical use, the separate methods need to be pipelined together. Recently, several studies attempt to solve



(a) Architecture with L stacked RACL layers.

(b) Details of a single RACL layer.

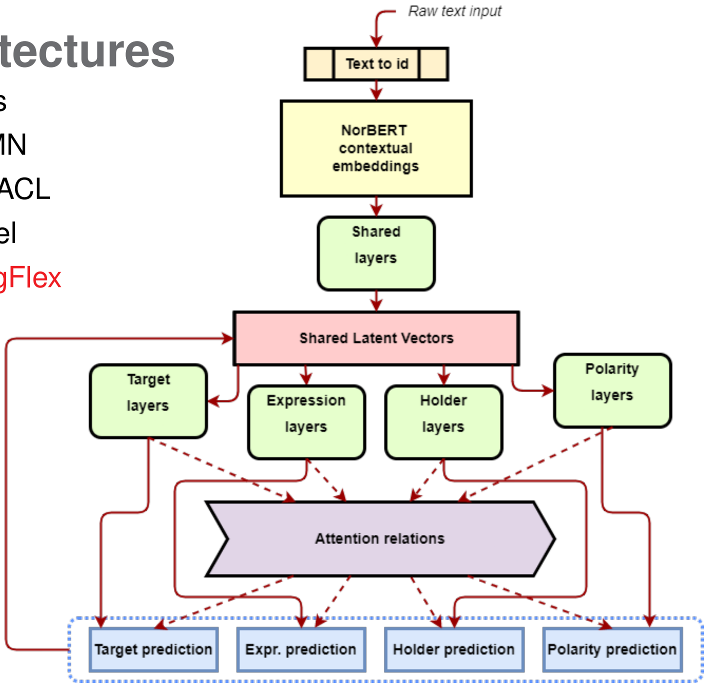
Architectures

Baselines

- IMN
- RACL

Our model

- FgFlex



Research Questions

Research questions

RQ1 Can we improve upon the current state-of-the-art multi-task learning models for extracting individual opinions from input sentences?

Research questions

RQ1 Can we improve upon the current state-of-the-art multi-task learning models for extracting individual opinions from input sentences?

RQ 2 What components are necessary for performance enhancement of fine-grained sequence-labelers on Norwegian data?

Research questions

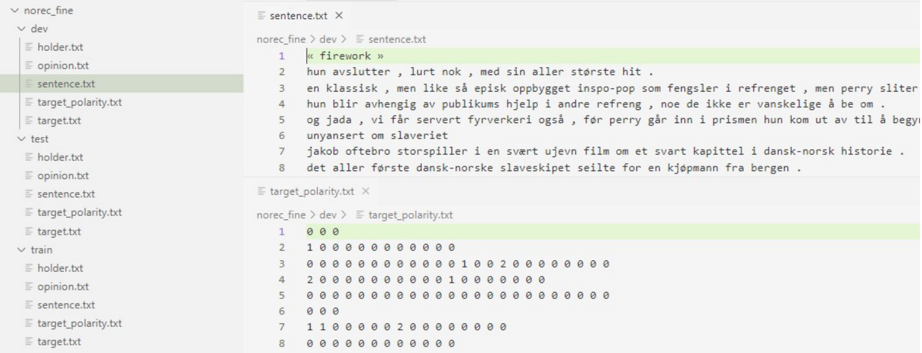
RQ1 Can we improve upon the current state-of-the-art multi-task learning models for extracting individual opinions from input sentences?

RQ 2 What components are necessary for performance enhancement of fine-grained sequence-labelers on Norwegian data?

Table of contents

First steps

- Build preprocessing pipeline



First steps

- Build preprocessing pipeline
- Test on baselines (original code)

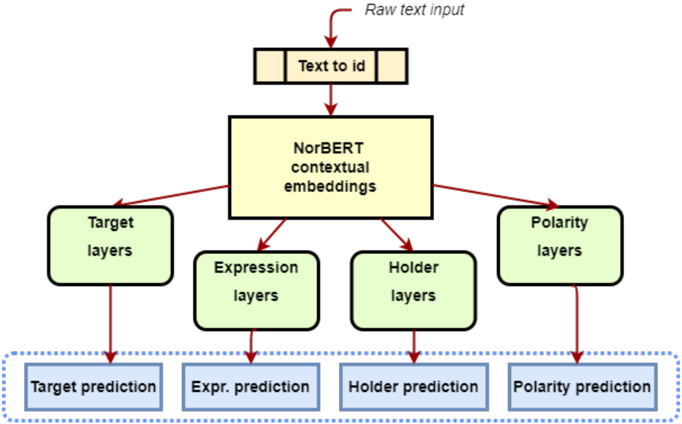
```
Validation results -- [Aspect f1]: 0.2375, [Opinion f1]: 0.1960, [Sentiment acc]: 0.7484, [Sentiment f1]: 0.3355, [Overall f1]: 0.1777
Test results -- [Aspect f1]: 0.2173, [Opinion f1]: 0.1862, [Sentiment acc]: 0.7476, [Sentiment f1]: 0.3759, [Overall f1]: 0.1624
Epoch 11, train: 322s
Validation results -- [Aspect f1]: 0.2930, [Opinion f1]: 0.2029, [Sentiment acc]: 0.7523, [Sentiment f1]: 0.3904, [Overall f1]: 0.2204
Test results -- [Aspect f1]: 0.2553, [Opinion f1]: 0.2016, [Sentiment acc]: 0.7955, [Sentiment f1]: 0.4537, [Overall f1]: 0.2031
Epoch 12, train: 319s
Validation results -- [Aspect f1]: 0.2945, [Opinion f1]: 0.2080, [Sentiment acc]: 0.7500, [Sentiment f1]: 0.4026, [Overall f1]: 0.2209
Test results -- [Aspect f1]: 0.2495, [Opinion f1]: 0.2034, [Sentiment acc]: 0.7923, [Sentiment f1]: 0.4379, [Overall f1]: 0.1977
Epoch 13, train: 35235s
Validation results -- [Aspect f1]: 0.2944, [Opinion f1]: 0.2090, [Sentiment acc]: 0.7934, [Sentiment f1]: 0.4027, [Overall f1]: 0.2336
Test results -- [Aspect f1]: 0.2537, [Opinion f1]: 0.2088, [Sentiment acc]: 0.7829, [Sentiment f1]: 0.4551, [Overall f1]: 0.1986
Epoch 14, train: 358s
Validation results -- [Aspect f1]: 0.3251, [Opinion f1]: 0.2013, [Sentiment acc]: 0.7662, [Sentiment f1]: 0.4162, [Overall f1]: 0.2491
Test results -- [Aspect f1]: 0.2797, [Opinion f1]: 0.2069, [Sentiment acc]: 0.7964, [Sentiment f1]: 0.4324, [Overall f1]: 0.2228
```

```
------------------------------------Iter5--------------------------------------
Train: final loss=157.814787, aspect loss=19.886210, opinion loss=101.958976, sentiment loss=16.078148, reg loss=524.332034, step=6474
Dev:   final loss=286.411440, aspect loss=65.815335, opinion loss=145.641024, sentiment loss=9.139039, reg loss=70.529764, step=6474
Dev:   aspect f1=0.4372, opinion f1=0.2575, sentiment acc==0.7575, sentiment f1==0.5081, ABSA f1==0.3312,
Test:  aspect f1=0.4330, opinion f1=0.2673, sentiment acc==0.7410, sentiment f1==0.5077, ABSA f1==0.3209,
Current Max Metrics Index : 0 Current Min Loss Index : 0 Epoch Time: 217m 36s
100%|          | 1079/1079 [3:37:39<00:00, 12.10s/it]
100%|          | 7/7 [20:12<00:00, 173.18s/it]
100%|          | 8/8 [24:18<00:00, 182.30s/it]

------------------------------------Iter6--------------------------------------
Train: final loss=112.135606, aspect loss=13.261519, opinion loss=72.868560, sentiment loss=12.738145, reg loss=586.240484, step=7553
Dev:   final loss=292.854676, aspect loss=70.058486, opinion loss=144.311758, sentiment loss=8.425107, reg loss=83.768051, step=7553
Dev:   aspect f1=0.4540, opinion f1=0.2644, sentiment acc==0.8301, sentiment f1==0.5251, ABSA f1==0.3769,
Test:  aspect f1=0.4257, opinion f1=0.2956, sentiment acc==0.7788, sentiment f1==0.4736, ABSA f1==0.3316,
```

First steps

- Build preprocessing pipeline
- Test on baselines (original code)
- Simple models



Decisions and motivations

- Rebuilding baselines in our codebase

```
class BertHead(torch.nn.Module): ⋯

class IMN(BertHead): ⋯

class RACL(IMN): ⋯

class FgFlex(BertHead): ⋯
```

Decisions and motivations

- Rebuilding baselines in our codebase
- Linear hyperparameter tuning

```
{
    "name": "fgflex-layers",

    "model_name": "fgflex",

    "shared_layers": [1,2,3,4,5],
    "expression_layers": [1,2,3,4],
    "polarity_layers": [1,2,3,4],
    "target_layers": [1,2,3,4],
```

Decisions and motivations

- Rebuilding baselines in our codebase
- Linear hyperparameter tuning
- Flexible architecture

FgFlex novelties

- relations between any subtask
- configurable convolutional filters
- activate/deactivate subtasks

Table of contents

Results

$$P_{absa} = \frac{TP_{target} + TP_{polarity}}{TP_{target} + TP_{polarity} + FP_{target} + FP_{polarity}} \quad (1)$$

$$R_{absa} = \frac{TP_{target} + TP_{polarity}}{TP_{target} + TP_{polarity} + FN_{target} + FN_{polarity}} \quad (2)$$

$$F1_{absa} = 2 \cdot \left(\frac{P_{absa} \cdot R_{absa}}{P_{absa} + R_{absa}} \right) \quad (3)$$

Model	Development data	Hold-out evaluation data
BertHead	0.4174	0.4152
FgsaLSTM	0.3877	0.3861
IMN	**0.4439**	**0.4438**
RACL	0.4201	0.4223
FgFlex	0.4129	0.4036

Hypothesis test

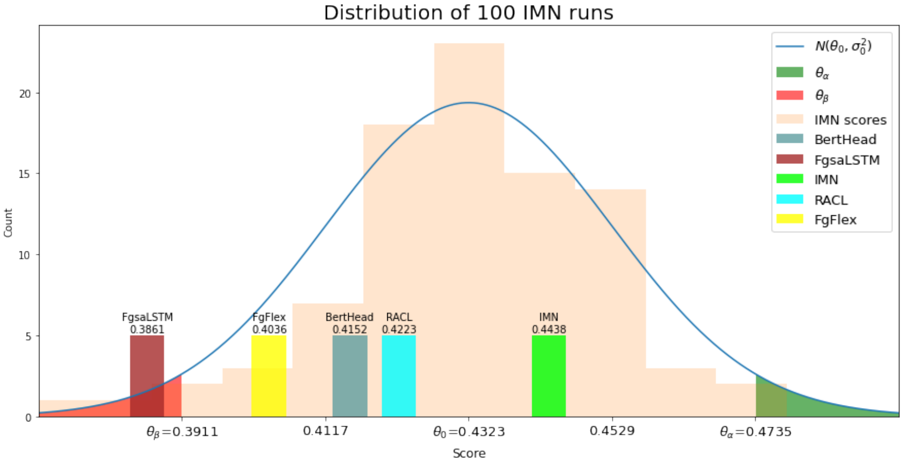


Table of contents

Future work

Bugs

- Sub-word evaluations
- Warm-up constant, opinion transmission, epoch count
- Others

Improvements

- Test coverage
- Reconstructing baselines
- Hyperparameter search

Directions for future work

- Metrics focused loss
- Auxiliary tasks
- More hyperparameter searching

References I

Barnes, J., Kurtz, R. et al.
'Structured Sentiment Analysis as Dependency Graph Parsing'.
Association for Computational Linguistics, 2021.

Chen, Z. and Qian, Tieyun.
'Relation-aware collaborative learning for unified aspect-based sentiment analysis'.
Association for Computational Linguistics,, 2020.

He, R. et al.
'An Interactive Multi-Task Learning Network for End- to-End Aspect-Based Sentiment Analysis'.
Association for Computational Linguistics, 2019.

Kutuzov, A. et al.
'Large-Scale Contextualised Language Modelling for Norwegian'.
Nordic Conference on Computational Linguistics, 2021.

References II

Velldal, E. et al.
'NoReC: The Norwegian Review Corpus'.
International Conference on Language Resources and Evaluation, 2018.

Øvrelid, L. et al.
' Fine-grained Sentiment Dataset for Norwegian'.
International Conference on Language Resources and Evaluation, 2020.

UiO : Department of Mathematics
University of Oslo

Per M.C. Halvorsen

FgFlex

A flexible, multitasking sequence-labeler for fine-grained sentiment analysis