Targeted Sentiment Analysis for Norwegian Language

2020-05-27

Ping-Han Hsieh

Targeted Sentiment Analysis

Sentiment analysis on *sentence* level

jeg elsker hunden min.
I love my dog.

But sometimes the sentence can be quite ambiguous

jeg hater katten hennes, men jeg elsker hunden min
I hate her cat, but I love my dog.

For finer-grained sentiment analysis, we want to know both the *target* and *polarity* of the sentiment

jeg hater katten hennes, men jeg elsker hunden min.

I hate her cat, but I love my dog.

NoReC$_{fine}$ (1)

Fine-grained sentiment analysis dataset from *Norweigan Review Corpus*

1. comments and reviews from video games, movies, products, restaurants, etc.
2. annotated with holders, targets and polar expression

Beginning-Inside-Outside (BIO) format

jeg	hater	katten	hennes	,	men	jeg	elsker	hunden	min
O	O	B-targ-Neg	I-targ-Neg	O	O	O	O	B-targ-Neg	I-targ-Neg

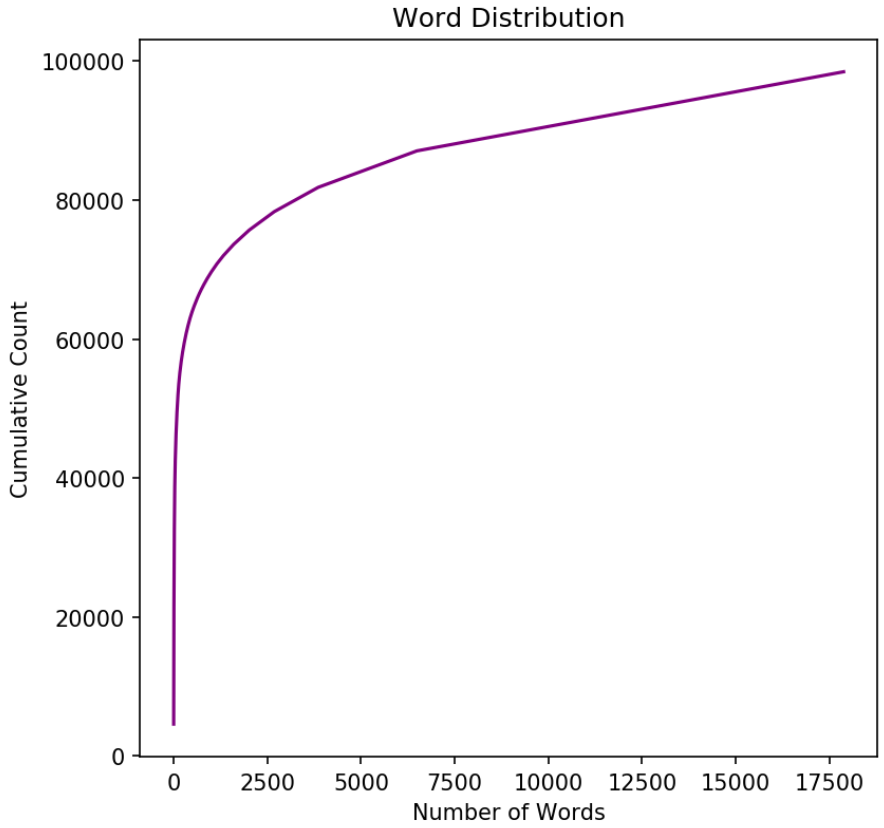
Predefined training, development and testing dataset

	# Examples				
	Train	Dev.	Test	Total	Avg. len.
Sents.	6145	1184	930	8259	16.8
Targets	4458	832	709	5999	2.0

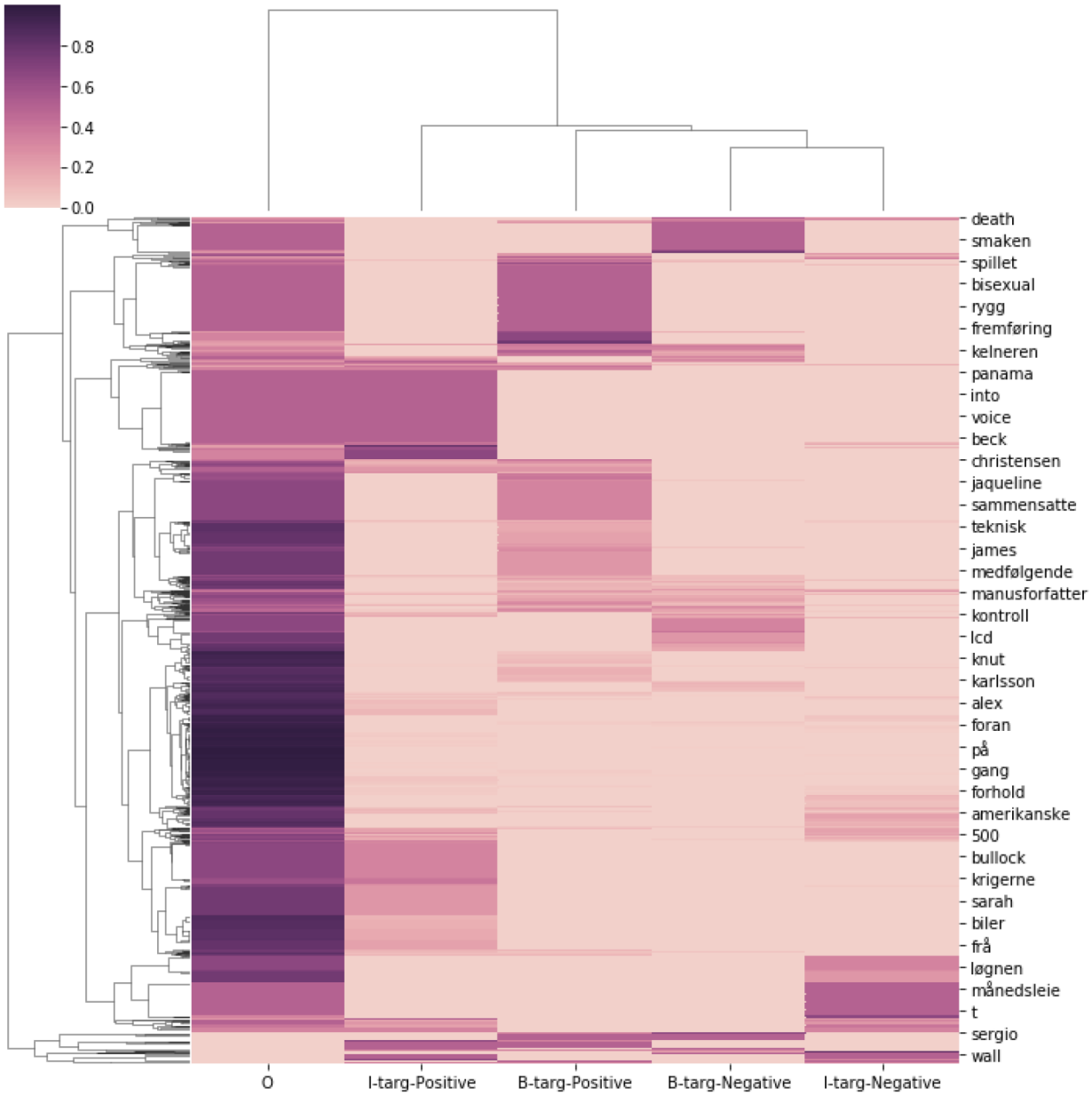
	Train	Dev	Test
O	91690	18335	14424
B-targ-Pos	2244	432	365
B-targ-Neg	1093	195	144
I-targ-Pos	2338	435	347
I-targ-Neg	1093	206	116

Øvrelid et al. 2019

NoReC$_{fine}$ (2)



* Most of the words have low frequency in the data



Neural Network Settings

Continuous word embeddings

100-dimensional pretrained embeddings (515,788 Norwegian words from Wikipedia) (Bojanowski et al. 2016)

Fixed hyperparameter settings

1. *word_dropout=0.01*
2. *Adam* optimizer with *learning_rate=0.05*
3. *batch_size=128*
4. *weighted cross entropy*
5. *train_epochs=20*

Explored hyperparameter settings

1. *number of hidden layers*
2. *number of neurons*

Explored network architectures

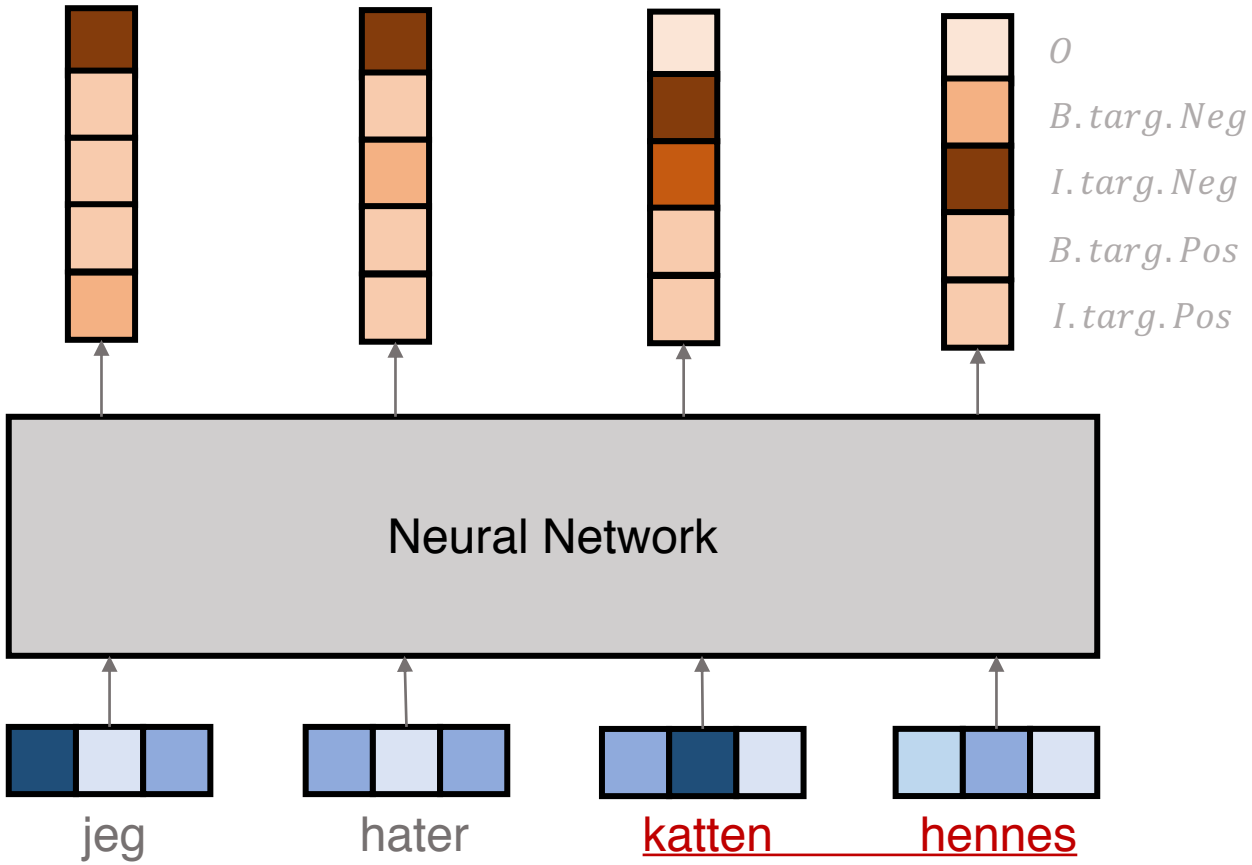
1. *Bidirectional LSTM (baseline)*
2. *Bidirectional GRU*
3. *Transformer* (Vaswani et al., 2017)

Evaluation

Proportional F1 (assigns **precision** and **recall** as the ratio of overlap with the predicted and gold span respectively)

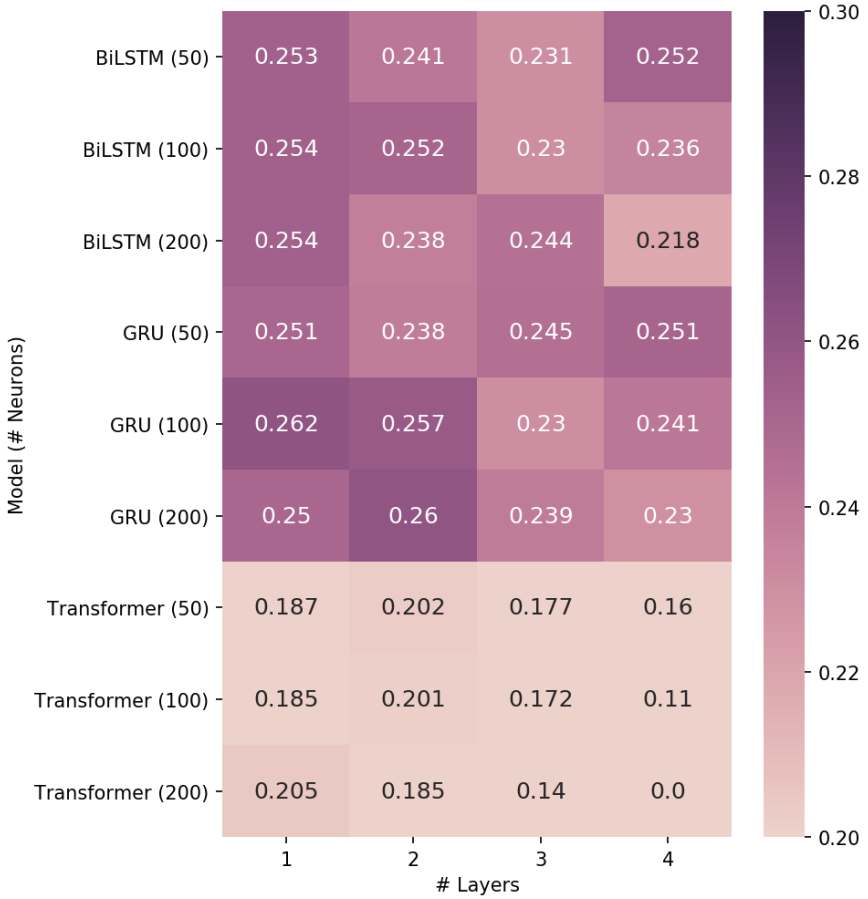
Modelling Strategies (1)

Collapsed Modelling



Experimental Results (1)

Collapsed Modelling



Baseline

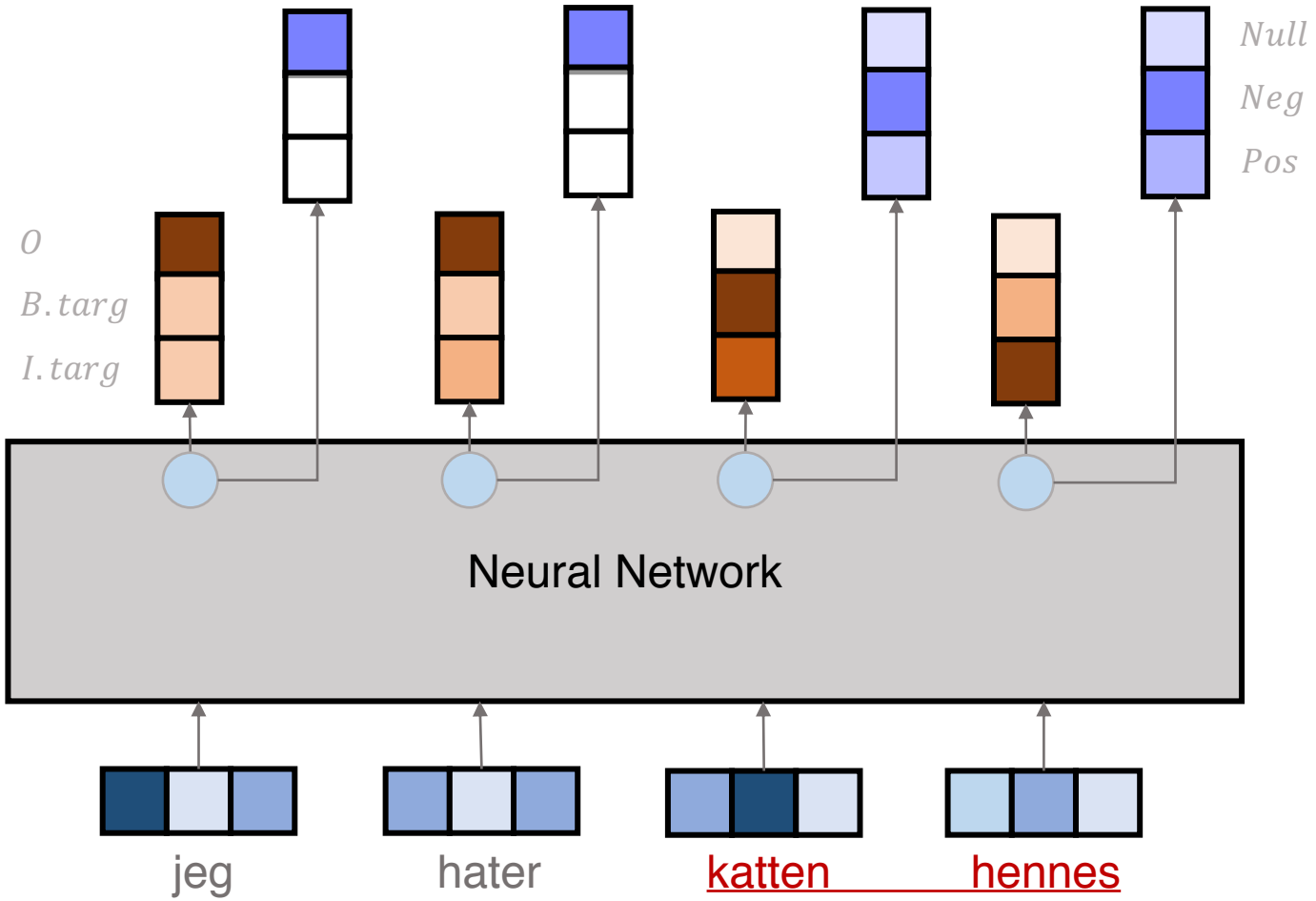
	Precision	Recall	Count
O	0.96	0.92	18335
B-targ-Pos	0.23	0.41	432
B-targ-Neg	0.11	0.15	195
I-targ-Pos	0.23	0.44	435
I-targ-Neg	0.09	0.09	206

GRU with 100 neurons in 1 hidden layer

	Precision	Recall	Count
O	0.97	0.92	18335
B-targ-Pos	0.22	0.49	432
B-targ-Neg	0.11	0.18	195
I-targ-Pos	0.27	0.37	435
I-targ-Neg	0.13	0.16	206

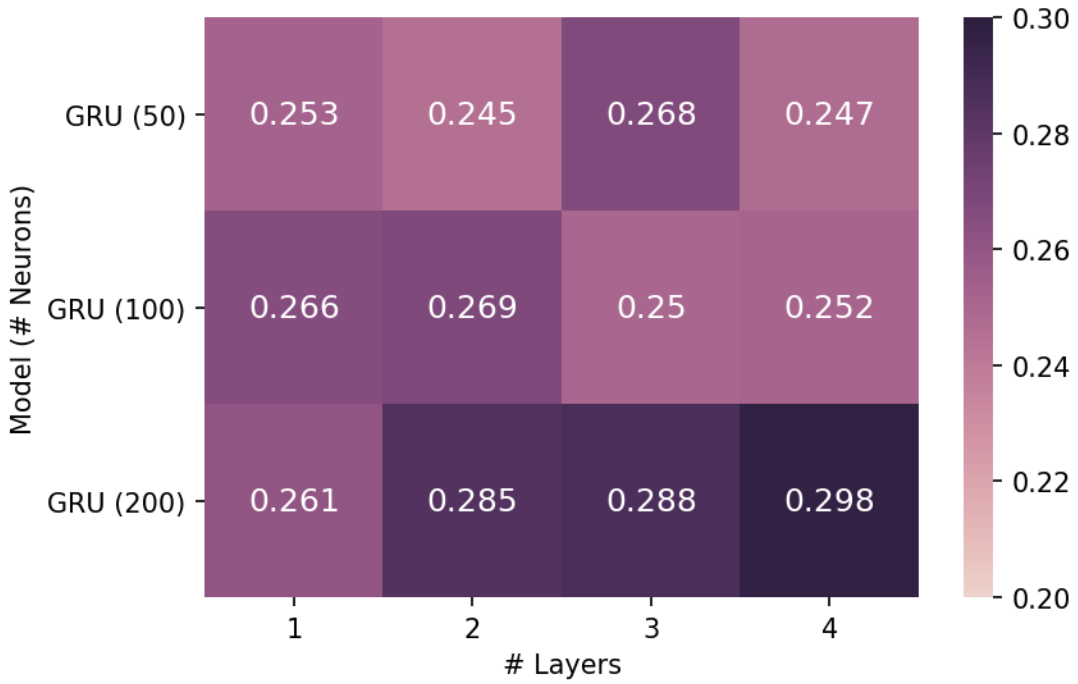
Modelling Strategies (2)

Joint Modelling



Experimental Results (2)

Joint Modelling



Baseline

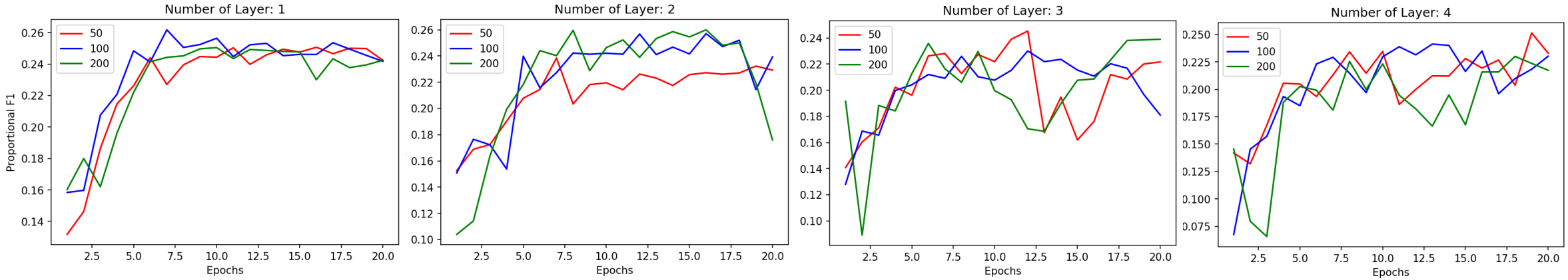
	Precision	Recall	Count
O	0.96	0.92	18335
B-targ-Pos	0.23	0.41	432
B-targ-Neg	0.11	0.15	195
I-targ-Pos	0.23	0.44	435
I-targ-Neg	0.09	0.09	206

GRU with 200 neurons in 4 hidden layers

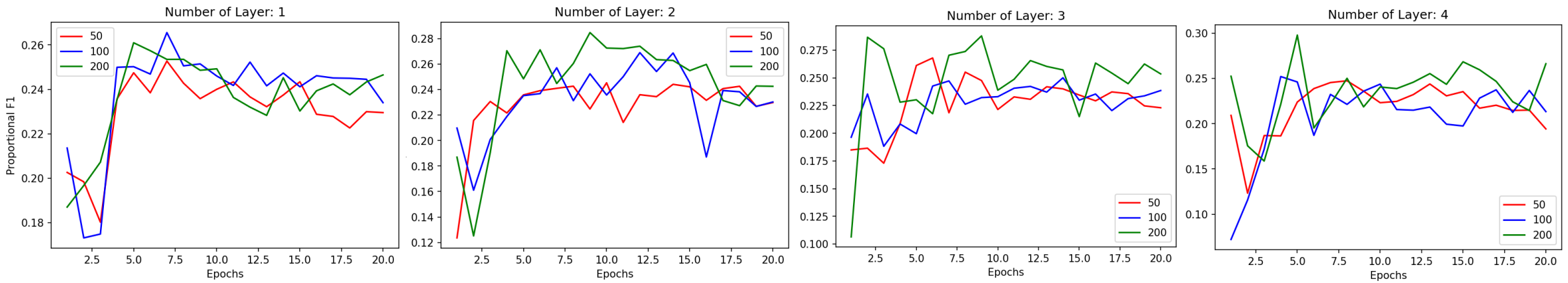
	Precision	Recall	Count
O	0.97	0.91	18335
B-targ-Pos	0.23	0.52	432
B-targ-Neg	0.18	0.24	195
I-targ-Pos	0.23	0.56	435
I-targ-Neg	0.21	0.19	206

Comparison for Convergence Speed

Collapse Modelling

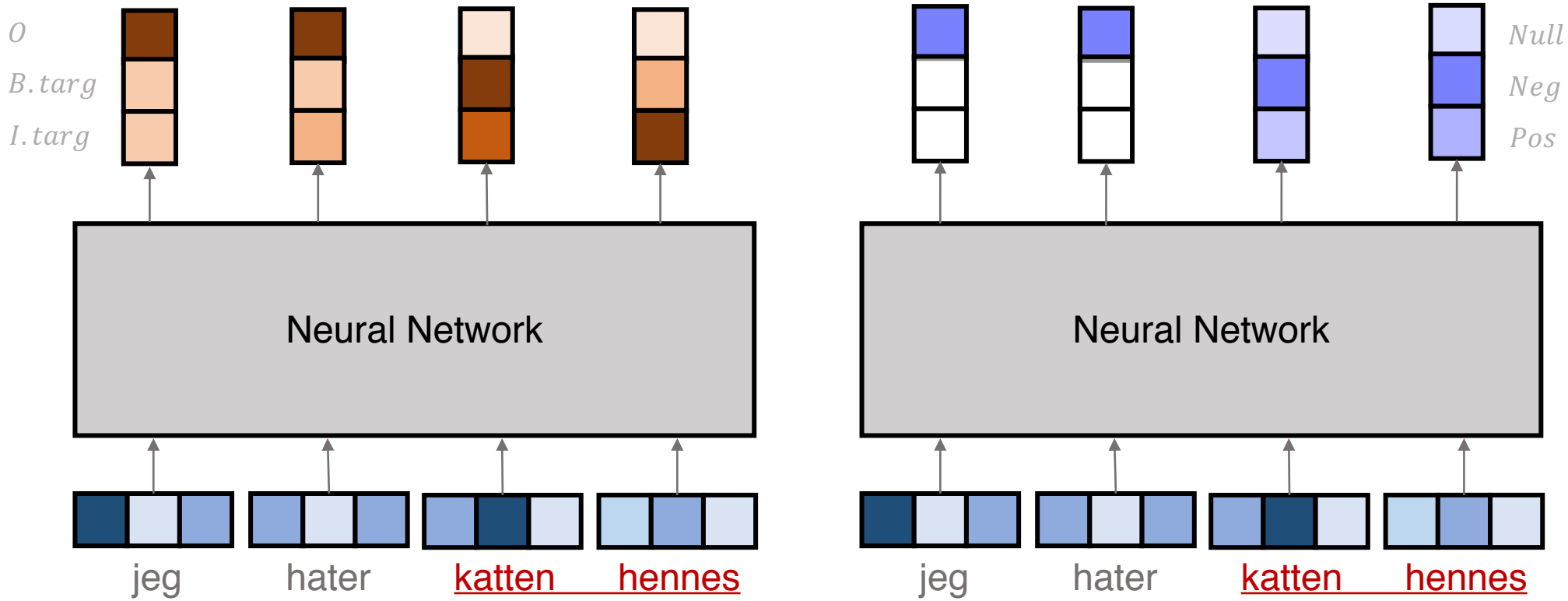


Joint Modelling



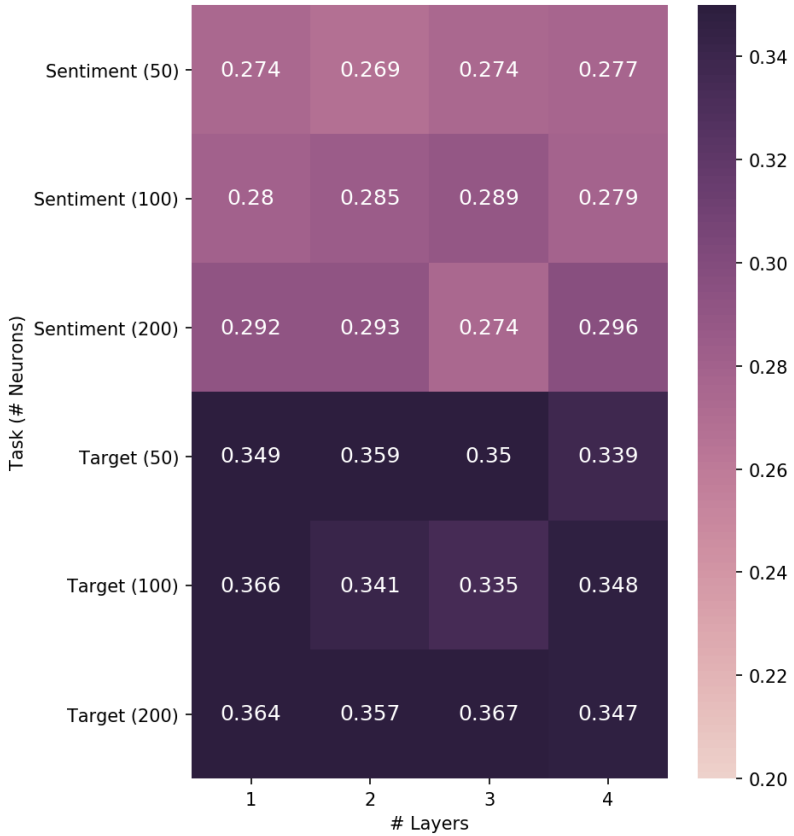
Modelling Strategies (3)

Pipeline Modelling



Experimental Results (3)

Pipeline Modelling * Combined Proportional F1 = **0.280**



Baseline

	Precision	Recall	Count
O	0.96	0.92	18335
B-targ-Pos	0.23	0.41	432
B-targ-Neg	0.11	0.15	195
I-targ-Pos	0.23	0.44	435
I-targ-Neg	0.09	0.09	206

GRU
with 200 neurons in 4 hidden layers (**Sentiment**)
with 200 neurons in 3 hidden layers (**Target**)

	Precision	Recall	Count
O	0.96	0.94	18335
B-targ-Pos	0.26	0.45	432
B-targ-Neg	0.18	0.17	195
I-targ-Pos	0.28	0.37	435
I-targ-Neg	0.13	0.09	206

Future Works and Conclusions

- Future work
  - Explore more hyperparameter settings (i.e. batch_size), gradient clipping, different word embedding, optimization and initialization algorithms.
  - Inspect why transformers perform comparatively worse (check implementation, try different numbers of multi-head attention, attention mechanisms and positional encoding)
  - Try pretrained model such as ELMo and BERT.
- Conclusions
  - Provide several experiment to explore the potential of neural networks and different modelling strategies.
  - Slight improvement over the baseline model on negative targeted entities.

Thanks