ACL 2021

Natural language processing

DefSent: Sentence Embeddings using Definition Sentences

Hayato Tsukagoshi, Ryohei Sasano, Koichi Takeda

Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing

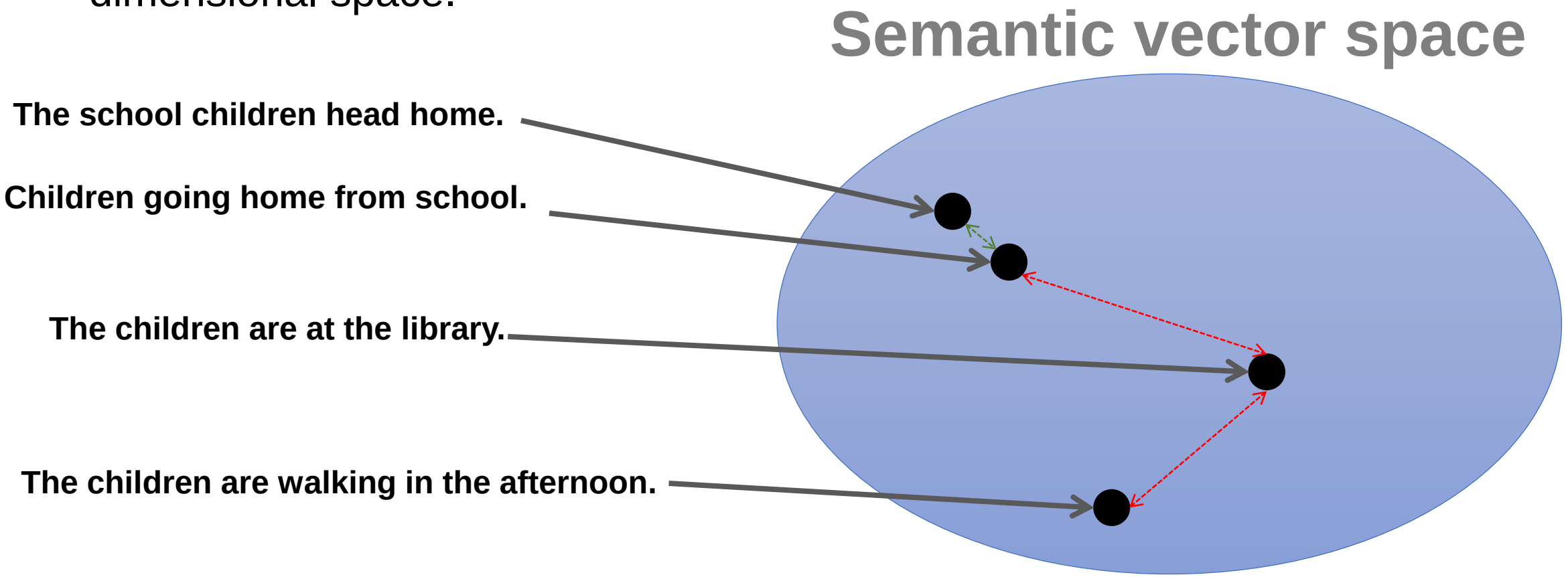
Instructor Mas. Nguyễn Thị Quý

- Introduction
- Method
- Experiment
- Conclusion

- Introduction
- Method
- Experiment
- Conclusion

What is sentence embedding

- Sentence embedding represent sentences as dense vectors in low dimensional space.



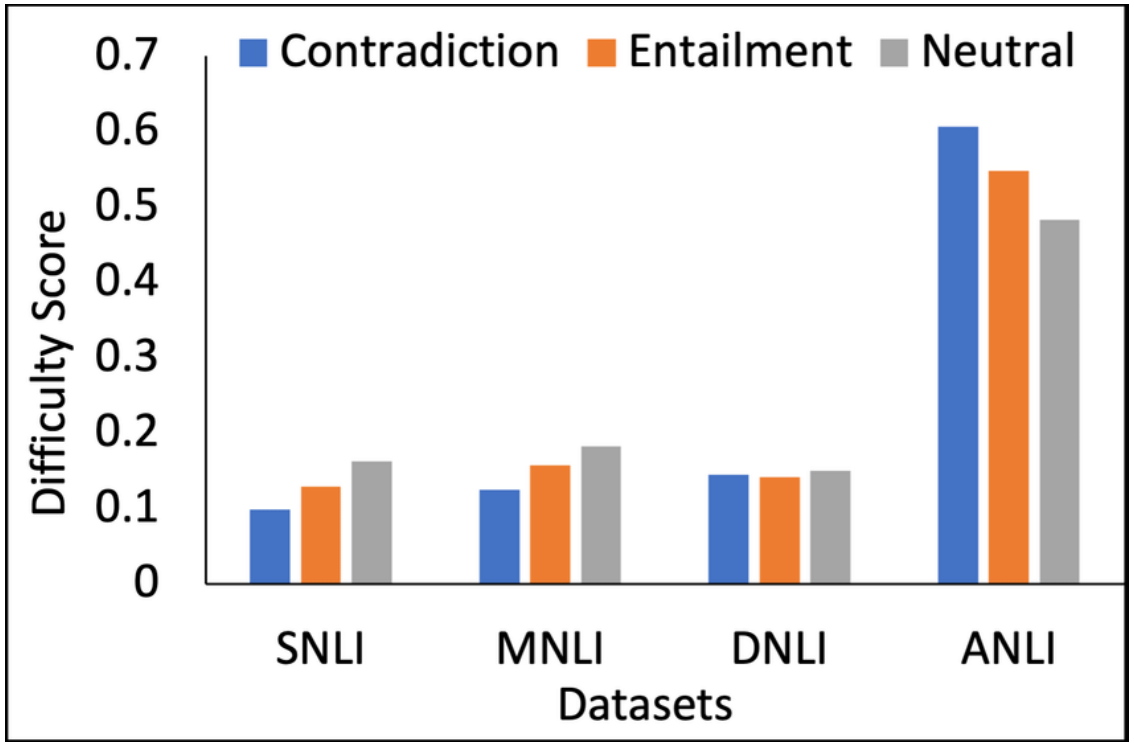
Introduction

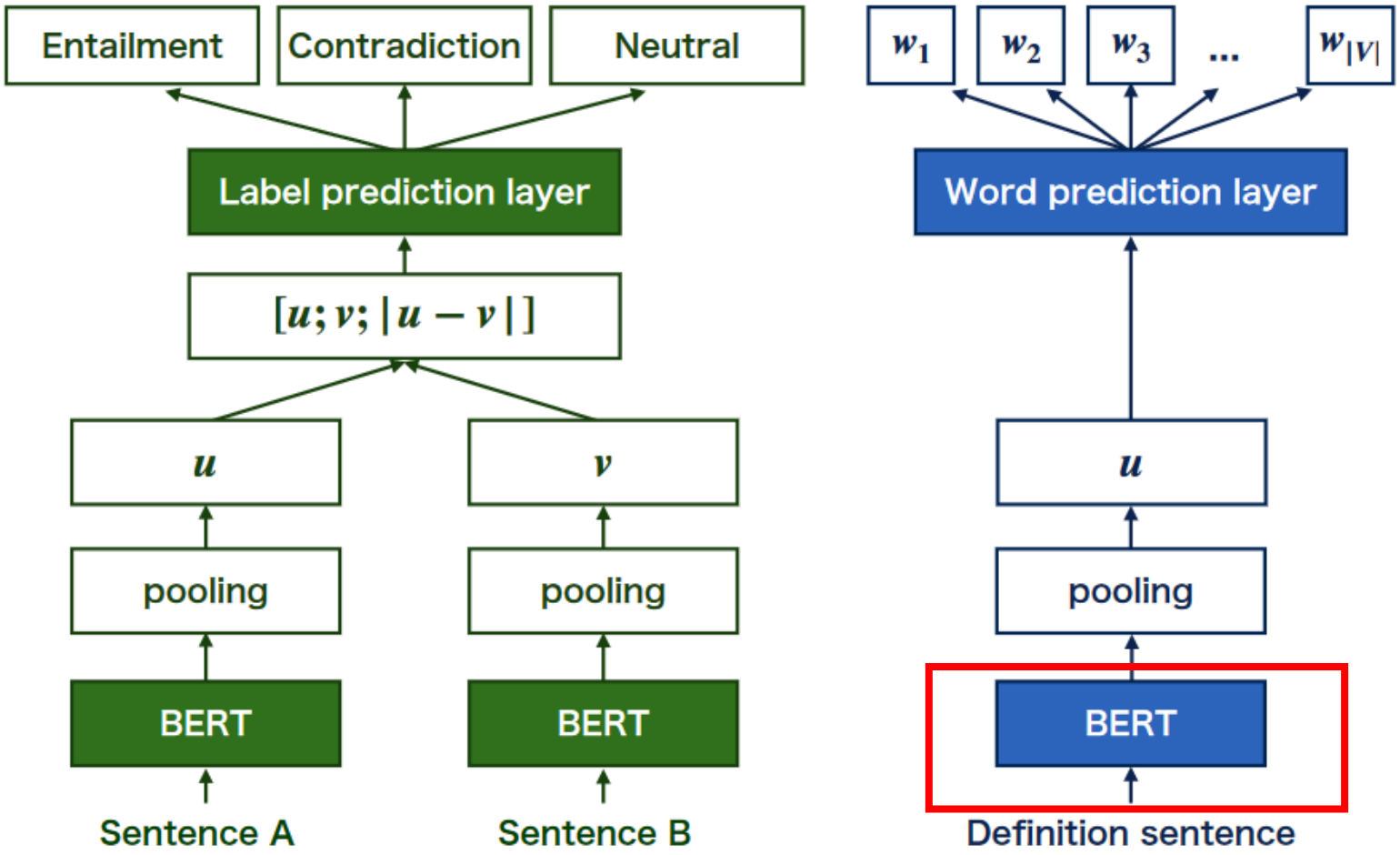
- Sentence embedding methods using natural language inference (NLI) datasets have been successfully applied to various tasks.
- These methods are only available for limited languages due to relying heavily on the large NLI datasets
- In this paper, They propose **DefSent**, a sentence embedding method that uses definition sentences from a word dictionary

→ This have more applicable uses

NLI dataset

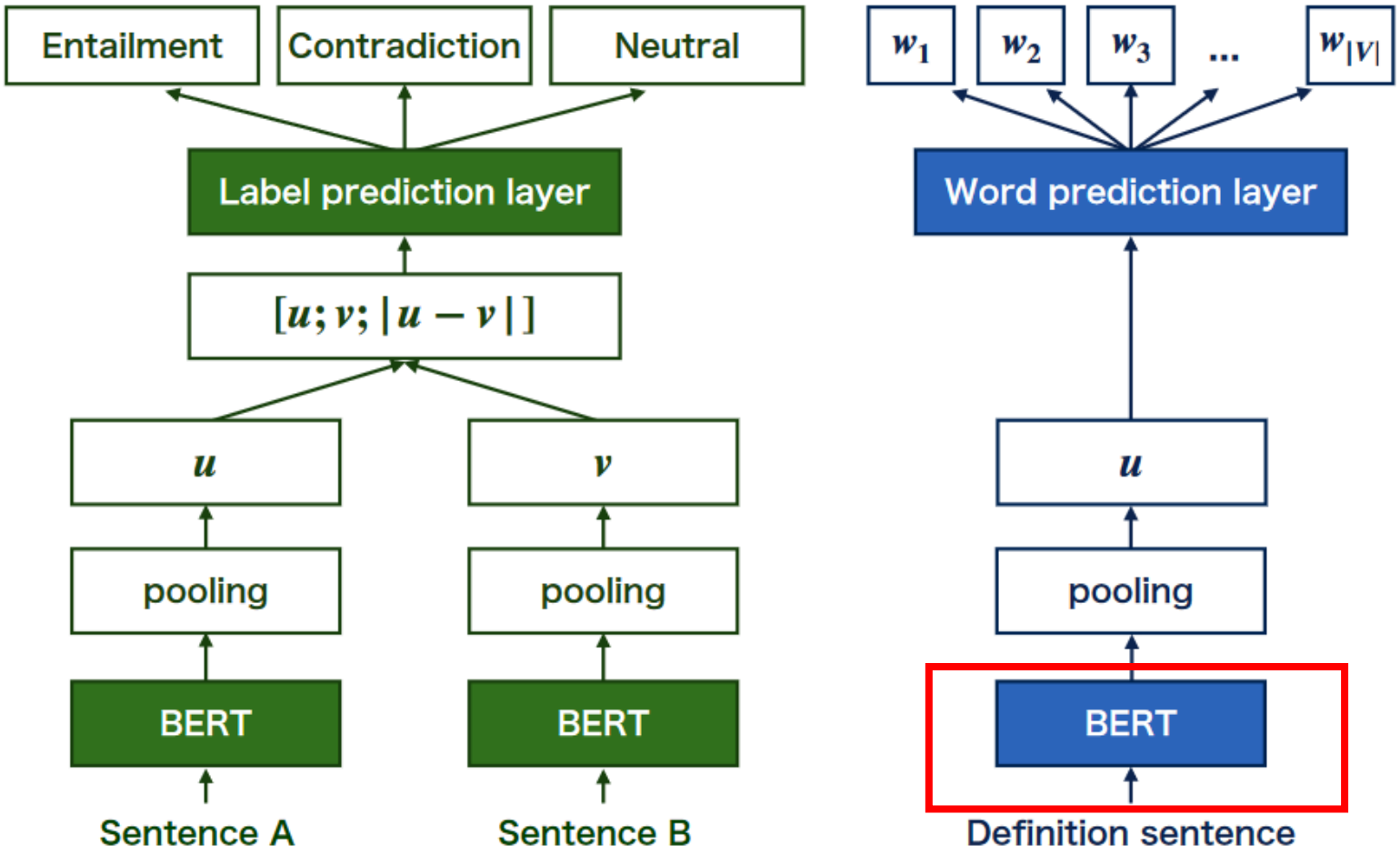
- Neutral
  a: Jon walked back to the town to the smithy.
  b: Jon traveled back to his hometown.
- Contradicts
  a: Tourist Information offices can be very helpful.
  b: Tourist Information offices are never of any help.
- Entails
  a: I'm confused.
  b: Not all of it is very clear to me.





Sentence-BERT (left) and DefSent (right).

- Introduction
- **Method**
- Experiment
- Conclusion



Sentence-BERT (left) and DefSent (right).

BERT

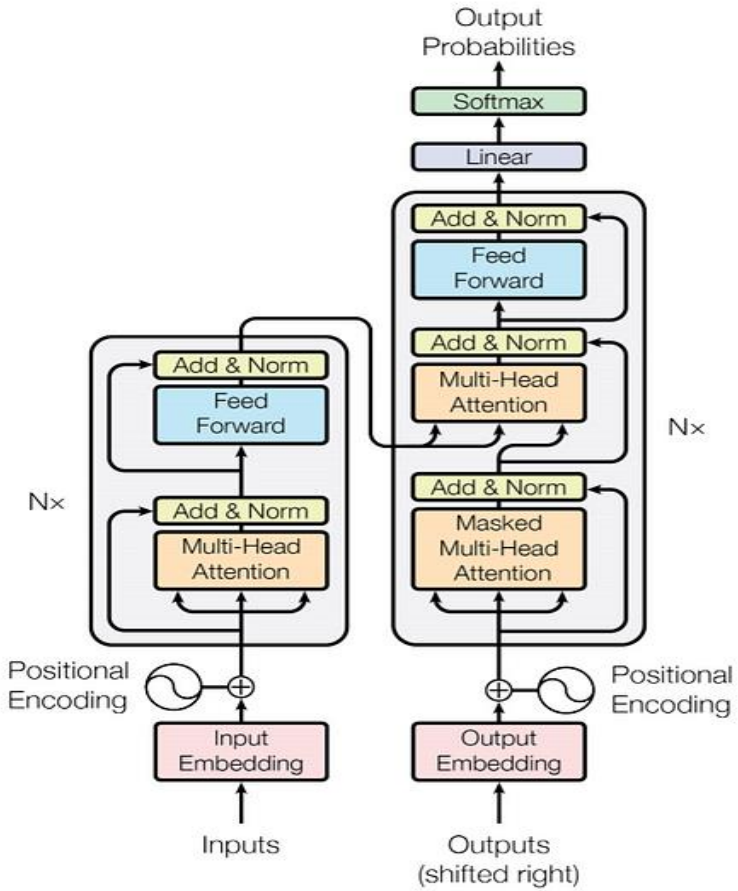


Figure 1: The Transformer - model architecture.

- BERT stand for: Bidirectional Encoder Representation from Transformer
- Transformers are made up of stacks of **transformer blocks**
- Simple linear layers feedforward networks, and **self-attention layers**

BERT

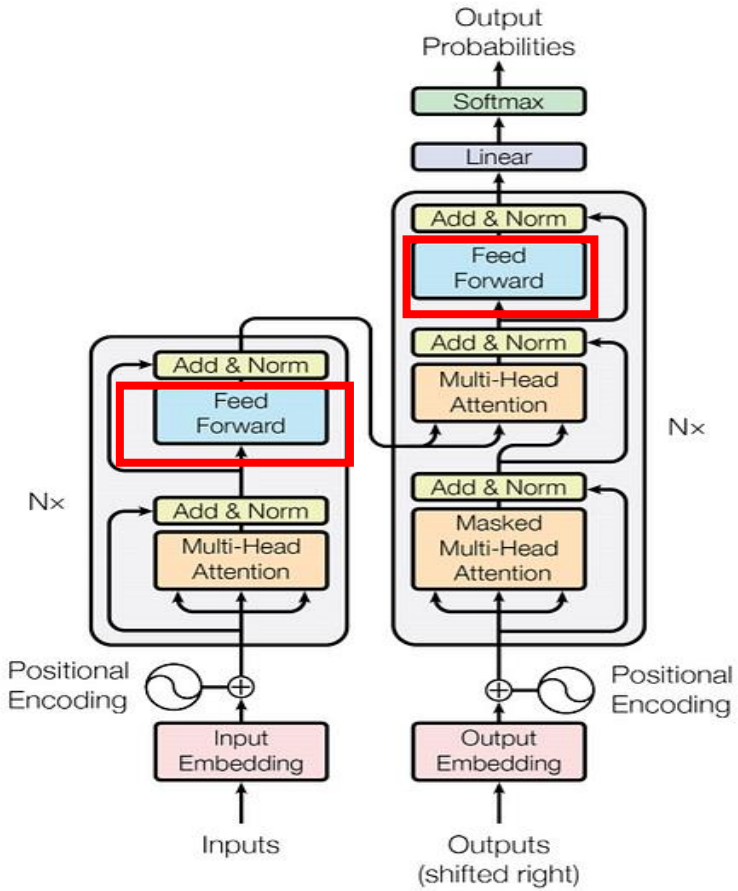


Figure 1: The Transformer - model architecture.

- BERT stand for: Bidirectional Encoder Representation from Transformer
- Transformers are made up of stacks of **transformer blocks**
- Simple linear layers feedforward networks, and **self-attention layers**

BERT

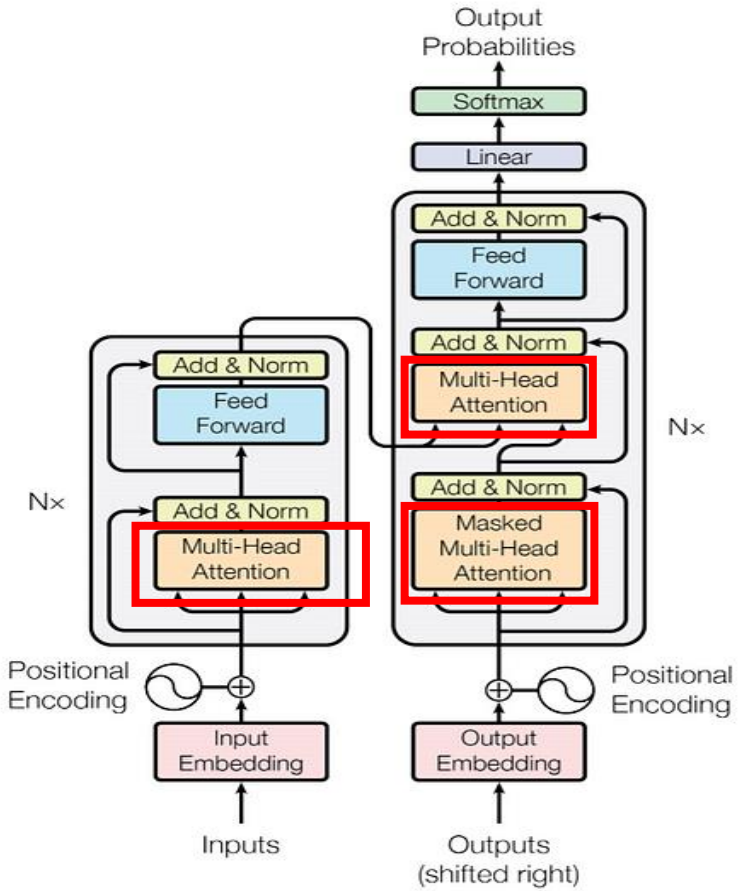


Figure 1: The Transformer - model architecture.

- BERT stand for: Bidirectional Encoder Representation from Transformer
- Transformers are made up of stacks of **transformer blocks**
- Simple linear layers feedforward networks, and **self-attention layers**

BERT

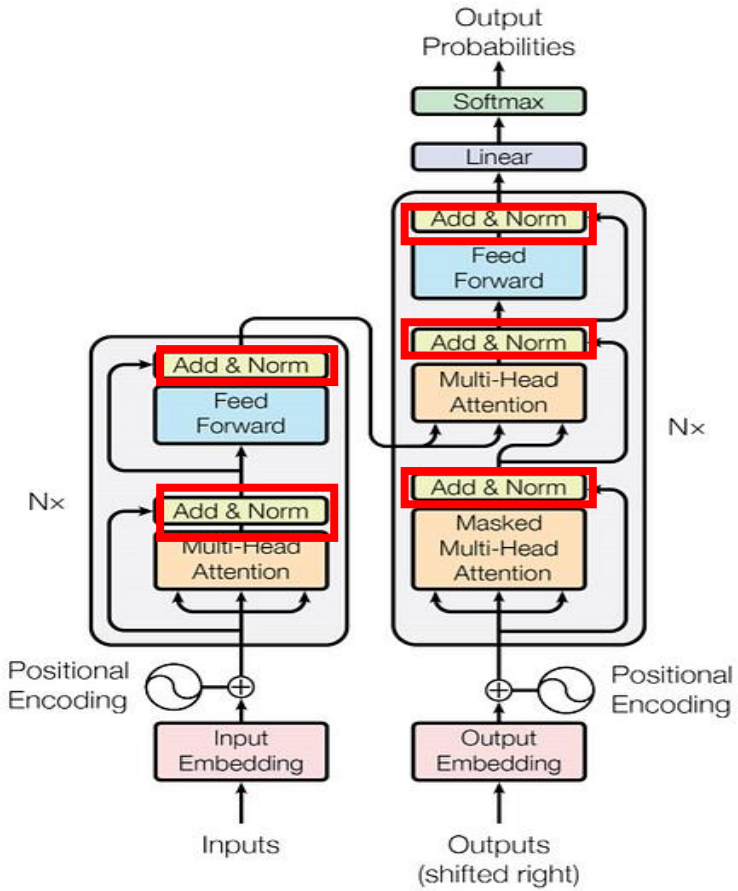


Figure 1: The Transformer - model architecture.

- BERT stand for: Bidirectional Encoder Representation from Transformer
- Transformers are made up of stacks of **transformer blocks**
- Simple linear layers feedforward networks, and **self-attention layers**

BERT

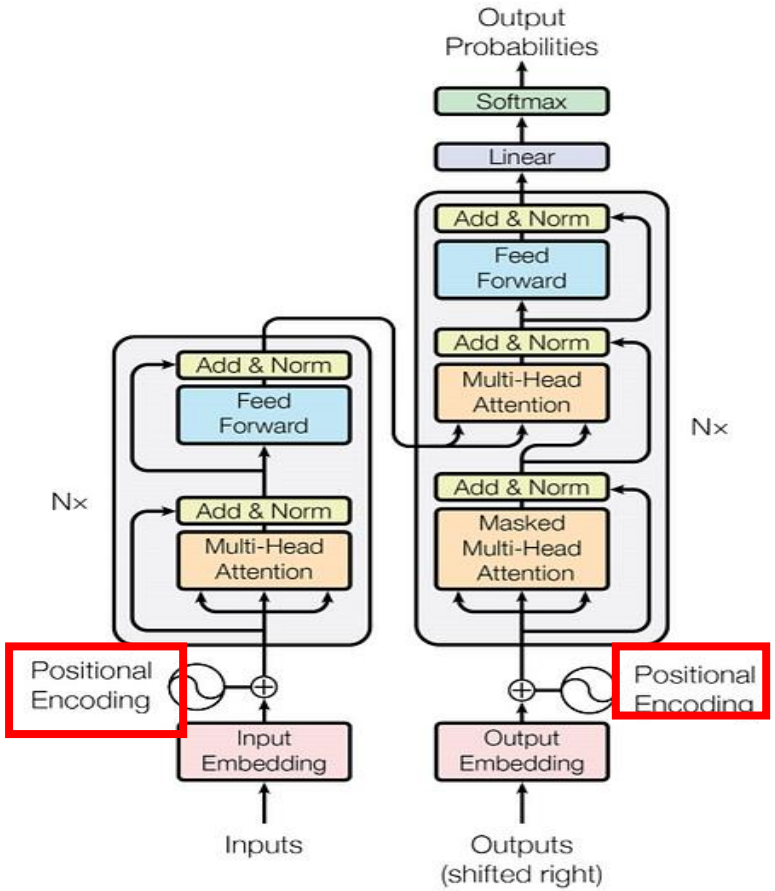
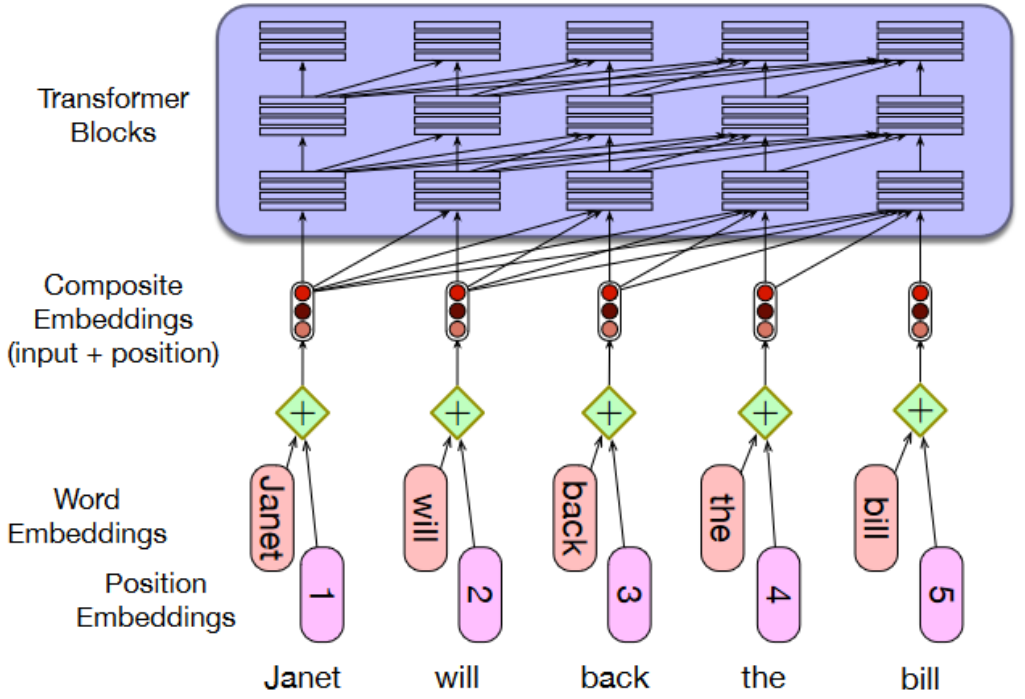
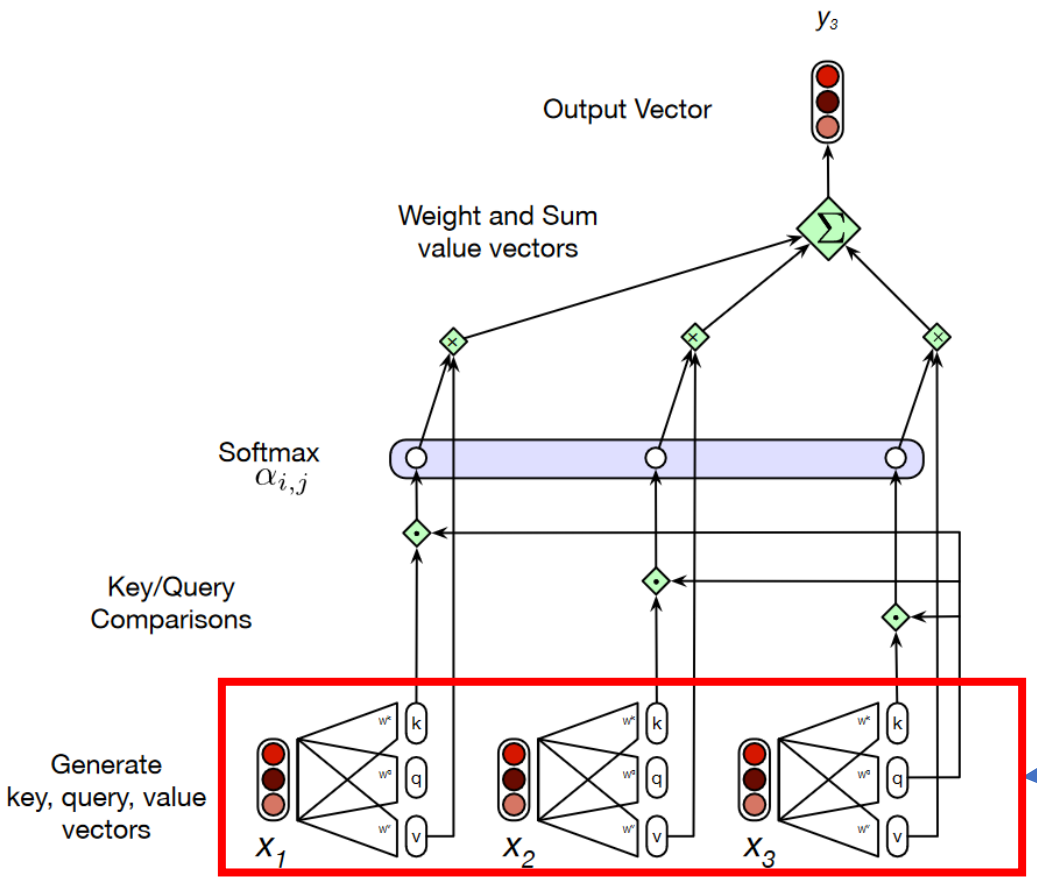


Figure 1: The Transformer - model architecture.



BERT



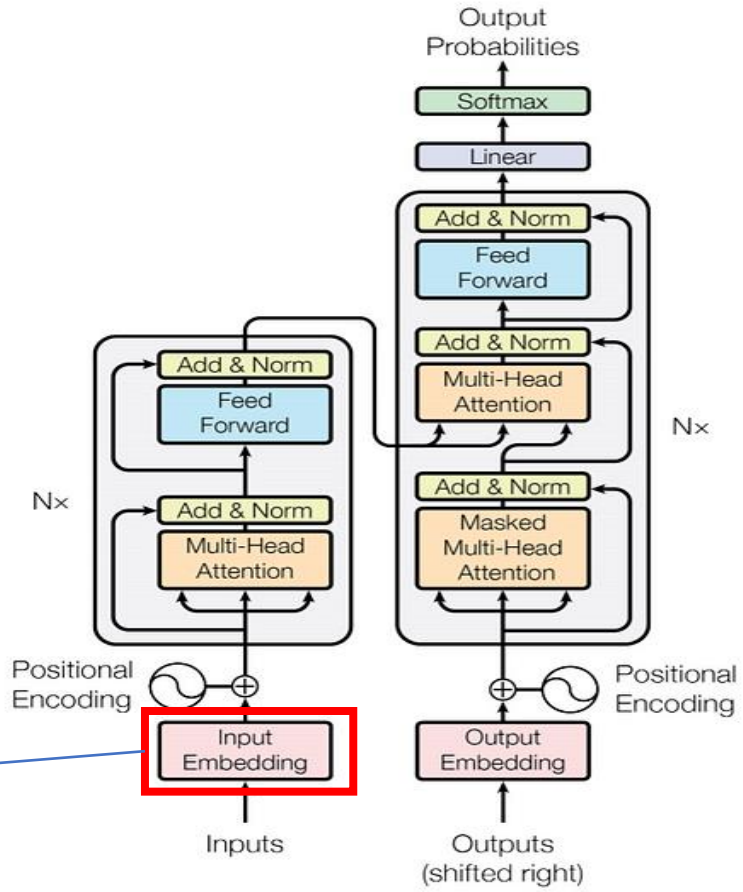
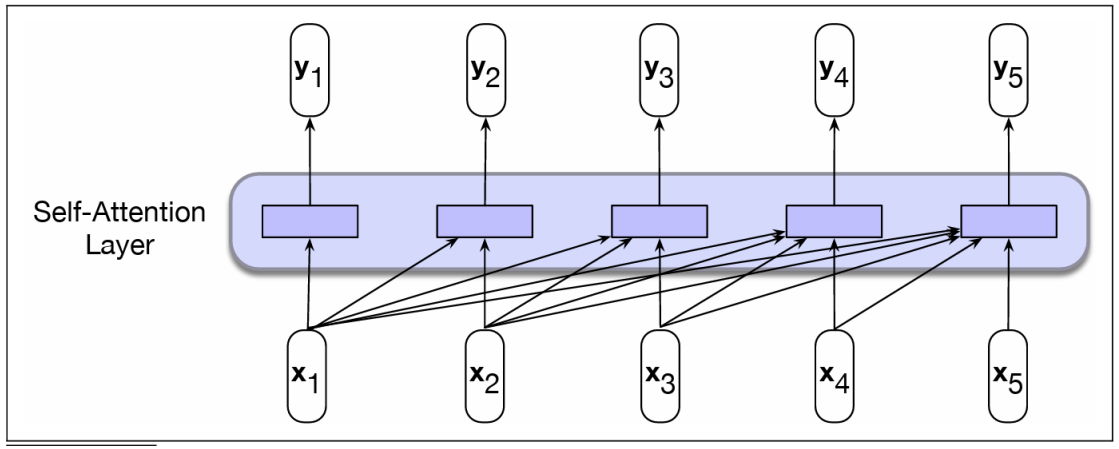
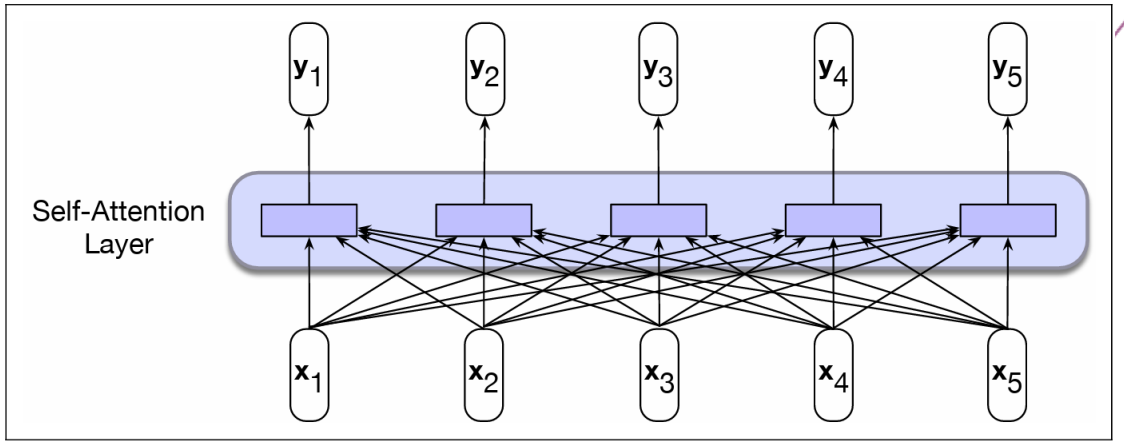


Figure 1: The Transformer - model architecture.

Self- Attention



Backward-looking self-attention model

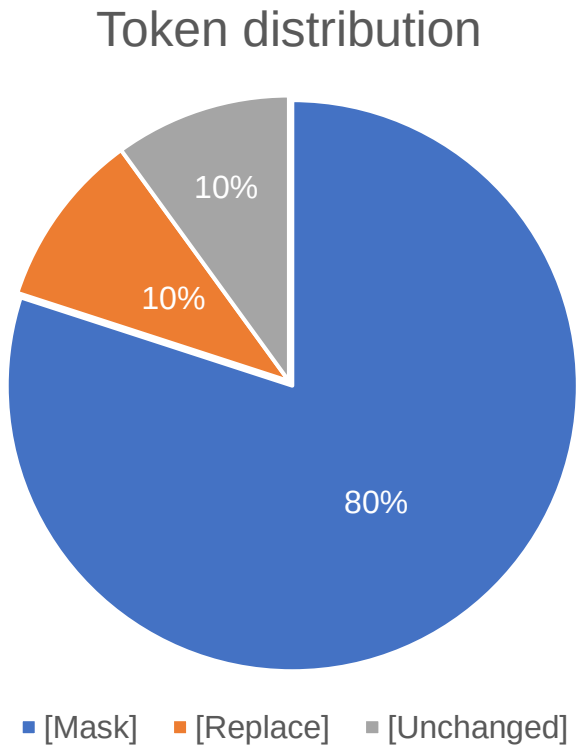


Bidirectional self-attention model

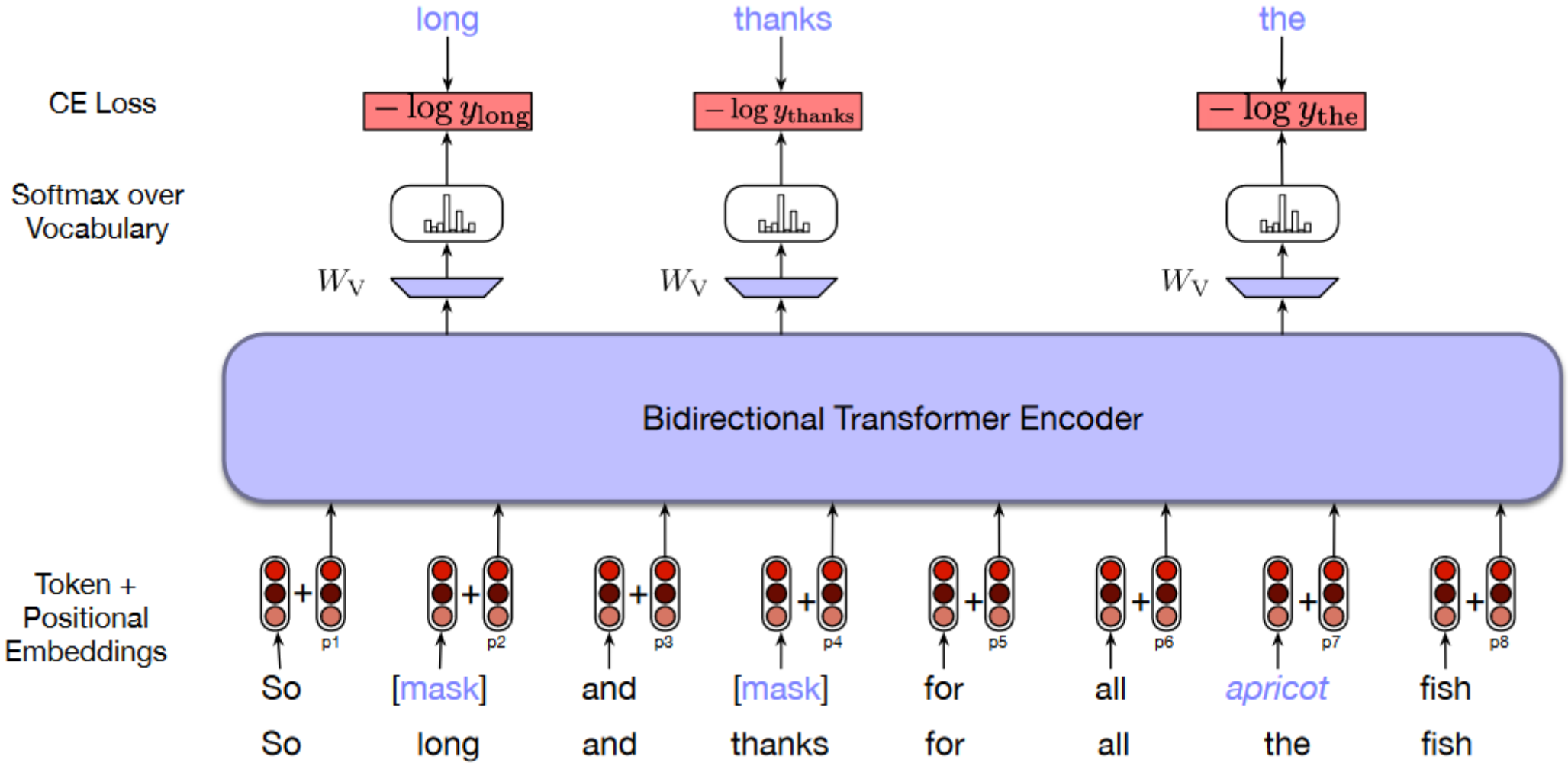
Compare with LSTMs : Can scalable, run paralel → Reduce time and utilize computation

Masked Language Model

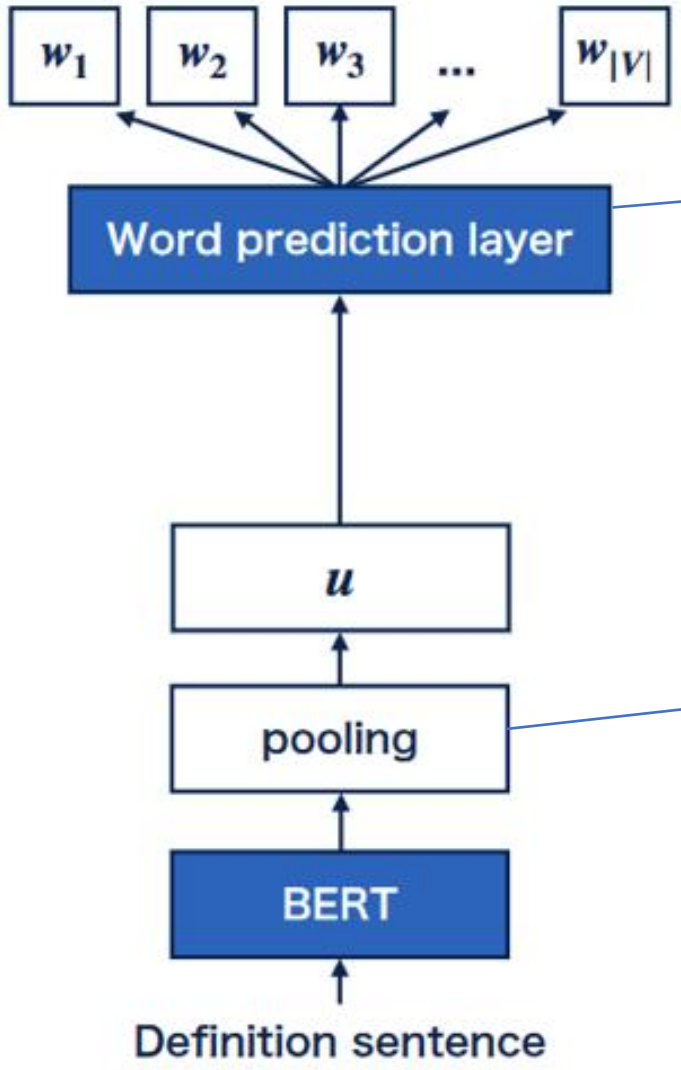
- **Masked Language Modeling (MLM)**: learns to perform a fill-in-the-blank task, technically called the **cloze task**
- In BERT, **15%** of the input tokens in a training sequence are sampled for learning



Masked Language Model



DefSent



Using Cross-entropy as loss function to train

CLS, Mean, Max

RoBERTa

1. Training the model longer, with bigger batches, over more data;
2. Removing the next sentence prediction objective;
3. Training on longer sequences
4. Dynamically changing the masking pattern applied to the training data.

Introduction

Method

Experiment

Conclusion

Word prediction

Oxford Dictionary dataset from Learning to Describe Unknown Phrases with Local and Global Contexts

Shonosuke Ishiwatari, Hiroaki Hayashi, Naoki Yoshinaga, Graham Neubig, Shoetsu Sato, Masashi Toyoda, Masaru Kitsuregawa

NAACL 2019

All	Words	Definitions	Avg. length
Train	29,413	97,759	9.921
Dev	3,677	12,127	9.874
Test	3,677	12,433	9.846
In BERT vocab.	**Words**	**Definitions**	**Avg. length**
Train	7,732	54,142	9.531
Dev	936	6,544	9.512
Test	979	6,930	9.551
In RoBERTa vocab.	**Words**	**Definitions**	**Avg. length**
Train	7,269	53,935	9.376
Dev	901	6,625	9.372
Test	925	6,945	9.41

Word prediction

Model	Pooling	MRR	Top1	Top3	Top10
BERT-base (no fine-tuning)	CLS	.0009	.0000	.0000	.0000
	Mean	.0132	.0001	.0043	.0242
	Max	.0327	.0157	.0320	.0626
BERT-base	CLS	.3200	.2079	.3670	.5418
	Mean	.3091	.1972	.3524	.5356
	Max	.2939	.1840	.3350	.5207
BERT-large	CLS	**.3587**	**.2388**	**.4139**	**.6011**
	Mean	.3286	.2091	.3792	.5723
	Max	.2925	.1814	.3356	.5194
RoBERTa-base	CLS	.3436	.2241	.3983	.5836
	Mean	.3365	.2170	.3906	.5783
	Max	.3072	.1941	.3523	.5386
RoBERTa-large	CLS	.3863	.2611	.4460	.6364
	Mean	**.3995**	**.2699**	**.4634**	**.6599**
	Max	.3175	.2015	.3646	.5543

0.299	**0.202**	**0.356**	**0.528**

0.317	**0.217**	**0.374**	**0.560**

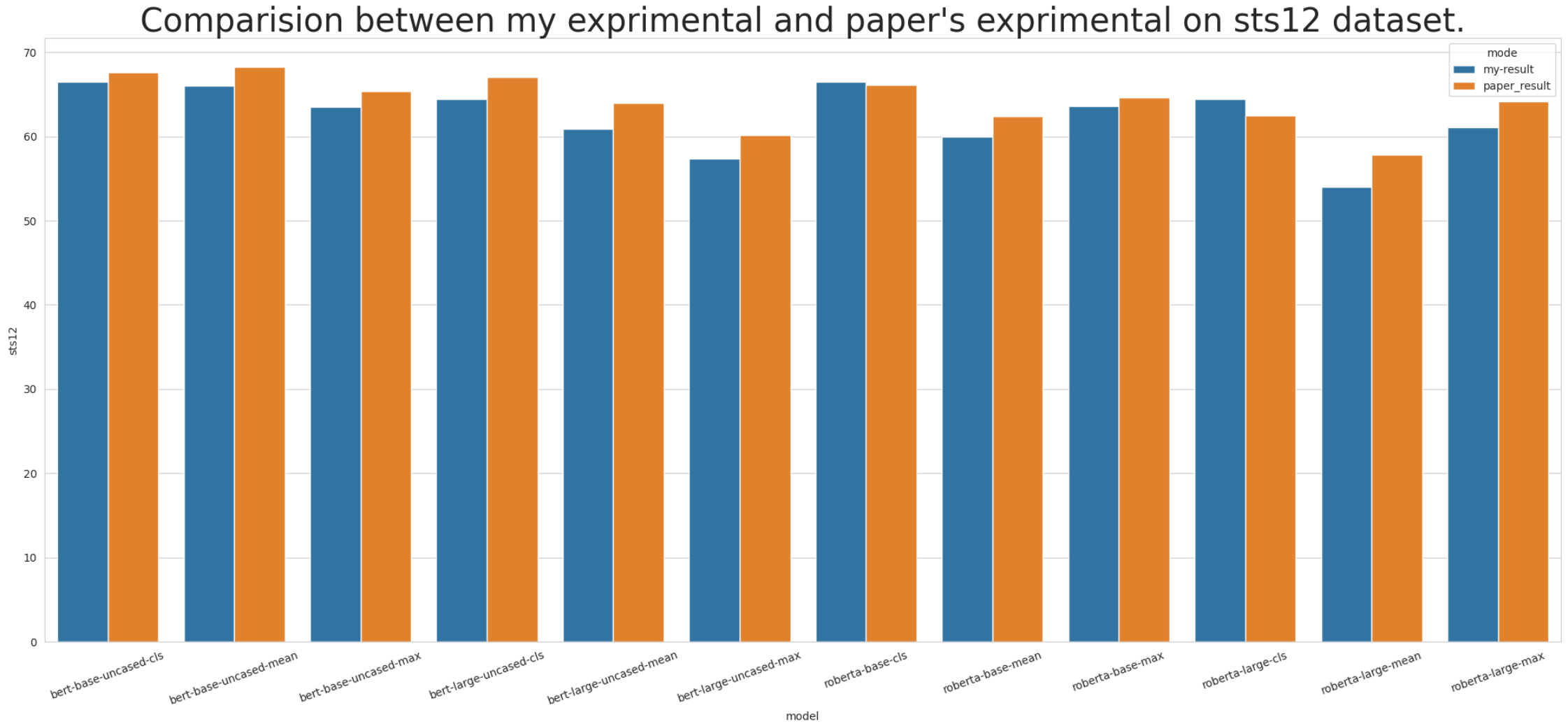
Semantic Textual Similarity(STS)

Model	STS12	STS13	STS14	STS15	STS16	STS-B	SICK-R	Avg.
Avg. GloVe embeddings (Pennington et al., 2014)	55.14	70.66	59.73	68.25	63.66	58.02	53.76	61.32
Avg. BERT embeddings	38.78	57.98	57.98	63.15	61.06	46.35	58.40	54.81
BERT CLS-vector	20.16	30.01	20.09	36.88	38.08	16.50	42.63	29.19
InferSent - Glove (Conneau et al., 2017)	52.86	66.75	62.15	72.77	66.87	68.03	65.65	65.01
Universal Sentence Encoder (Cer et al., 2018)	64.49	67.80	64.61	76.83	73.18	74.92	**76.69**	71.22
Sentence-BERT-base (Mean)	70.97	76.53	73.19	79.09	74.30	77.03	72.91	74.89
Sentence-BERT-large (Mean)	72.27	78.46	**74.90**	**80.99**	76.25	**79.23**	73.75	76.55
Sentence-RoBERTa-base (Mean)	71.54	72.49	70.80	78.74	73.69	77.77	74.46	74.21
Sentence-RoBERTa-large (Mean)	**74.53**	77.00	73.18	81.85	76.82	79.10	74.29	**76.68**
DefSent-BERT-base (CLS)	67.56	79.86	69.52	76.83	76.61	75.57	73.05	74.14
DefSent-BERT-large (CLS)	66.22	**82.07**	71.48	79.34	75.38	73.46	74.30	74.61
DefSent-RoBERTa-base (CLS)	65.55	80.84	71.87	78.77	**79.29**	78.13	74.92	75.62
DefSent-RoBERTa-large (Mean)	58.36	76.24	69.55	73.15	76.90	78.53	73.81	72.36

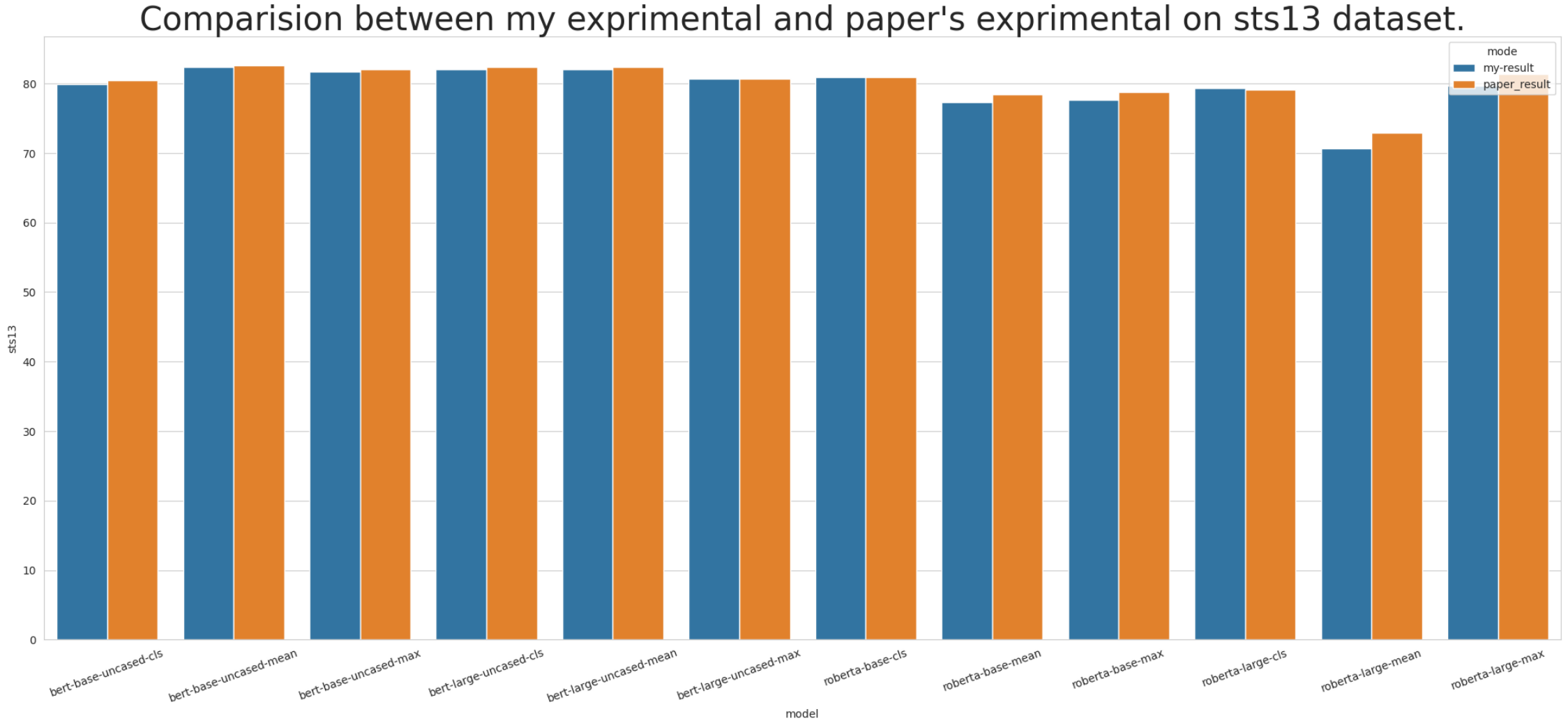
Semantic Textual Similarity(STS)

sts12	sts13	sts14	sts15	sts16	stsB	sickR	avg mode
66.448	79.946	68.042	76.941	75.009	76.743	71.589	73.531
65.996	82.346	72.274	78.666	75.315	78.881	70.791	74.896
63.519	81.728	71.966	78.216	74.727	77.722	70.644	74.075
64.446	82.001	69.375	80.390	74.265	75.214	72.743	74.062
60.849	82.076	73.009	80.055	76.625	79.514	73.213	75.049
57.379	80.675	72.122	77.572	74.722	77.552	71.989	73.144
66.431	80.958	71.950	80.120	78.598	80.348	74.349	76.108
59.968	77.366	68.844	76.478	76.815	78.992	72.453	72.988
63.560	77.626	68.012	77.534	78.222	80.356	73.770	74.154
64.413	79.298	72.893	75.870	76.984	80.204	74.776	74.920
54.017	70.686	66.853	73.120	73.636	80.527	73.906	70.392
61.099	79.687	71.361	76.857	78.685	80.693	72.933	74.473

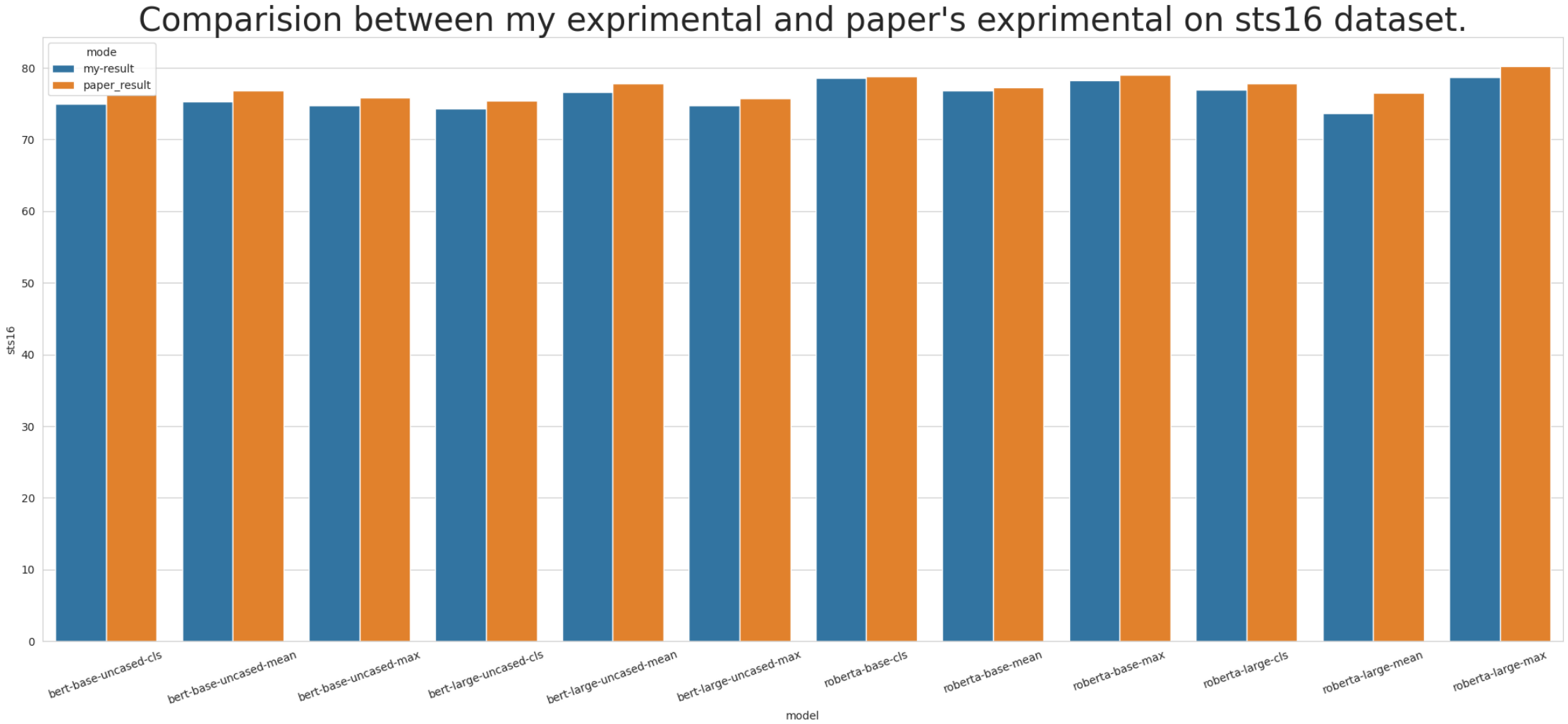
Semantic Textual Similarity(STS)



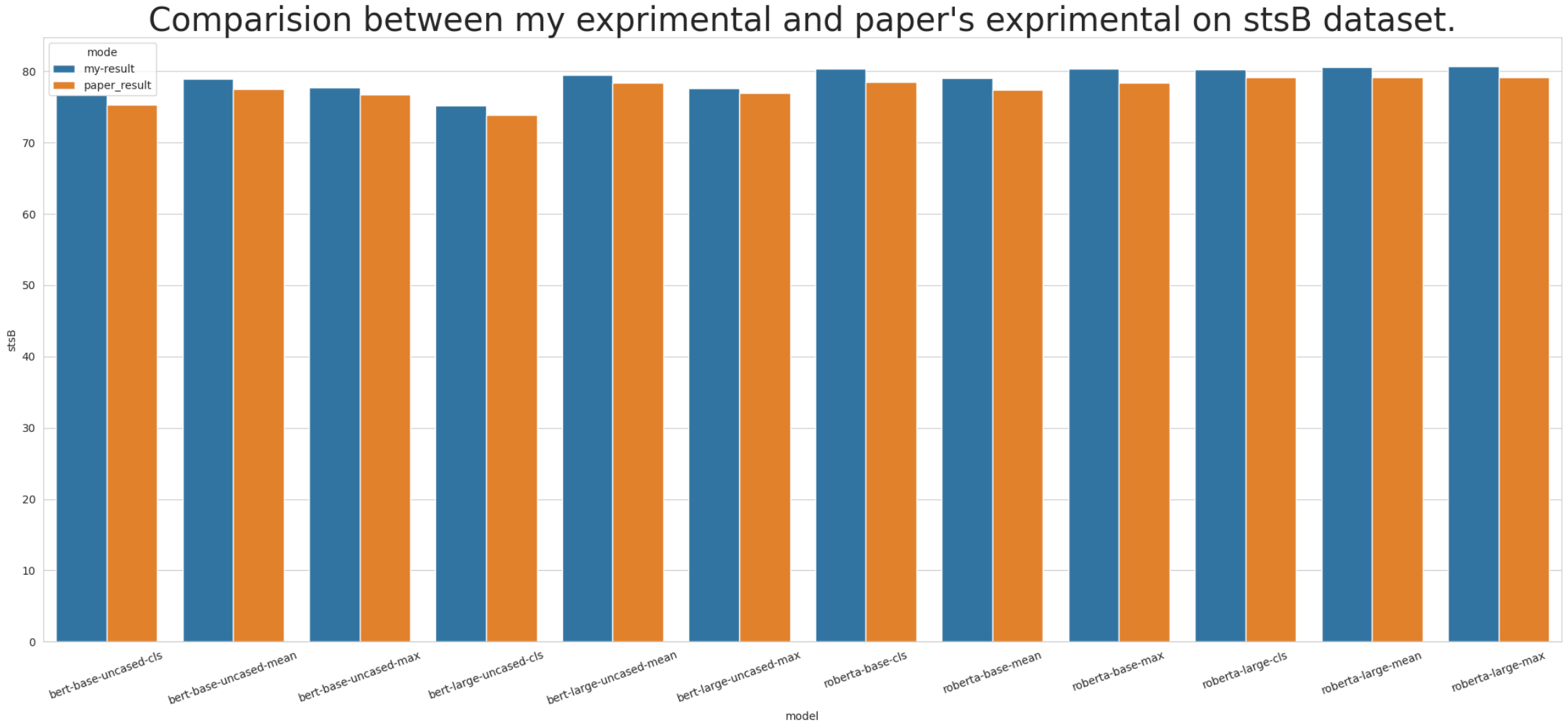
Semantic Textual Similarity(STS)



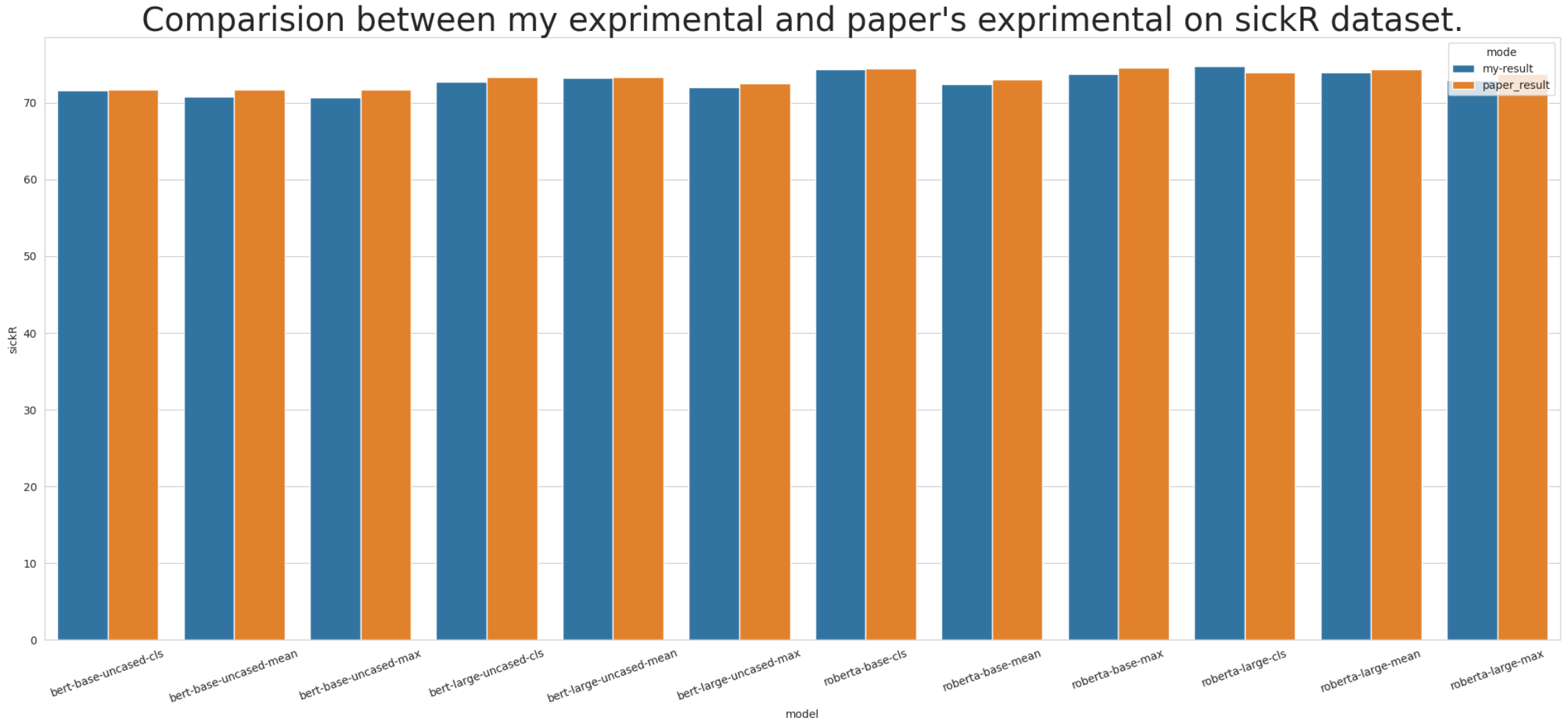
Semantic Textual Similarity(STS)



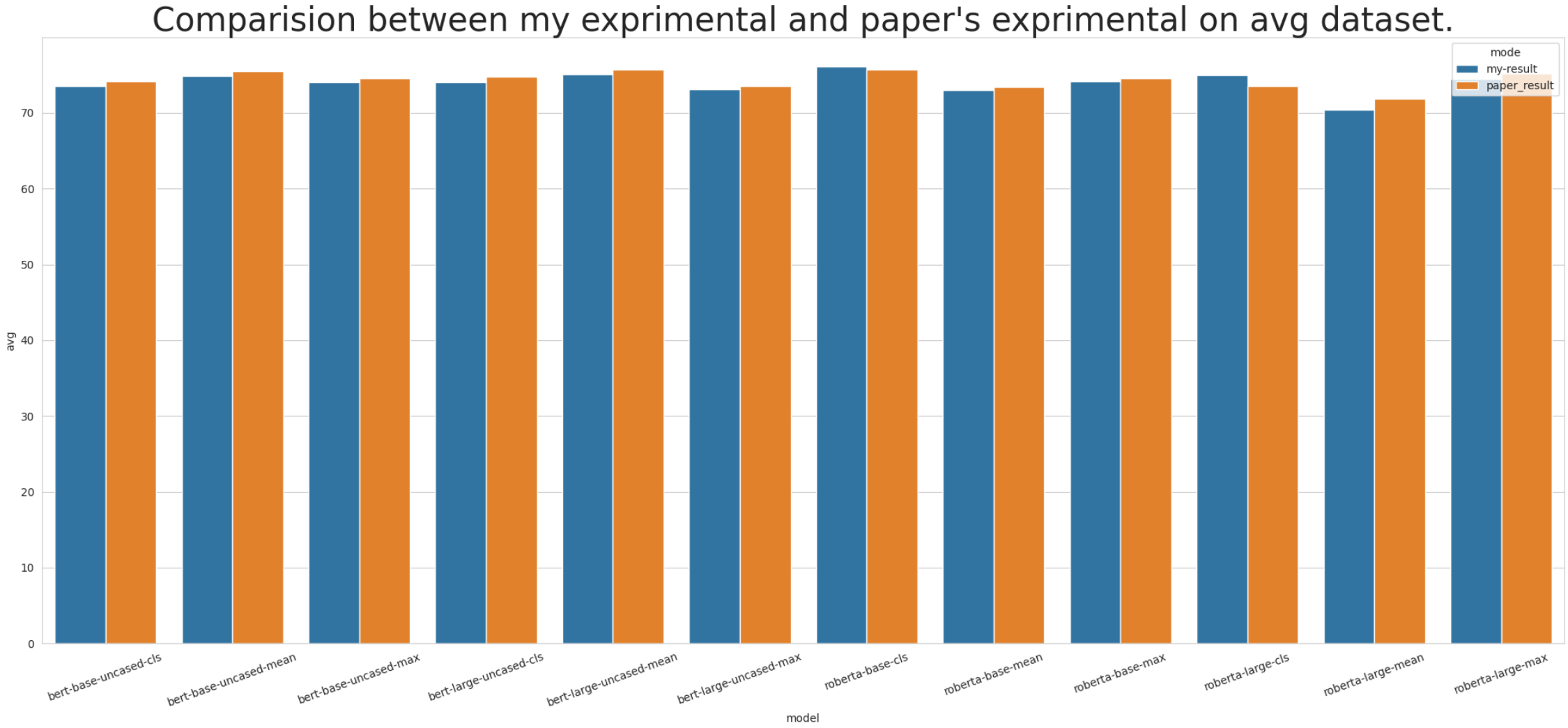
Semantic Textual Similarity(STS)



Semantic Textual Similarity(STS)



Semantic Textual Similarity(STS)



SentEval (Facebook researcher)

Task	Type	#train	#test	needs_train	set_classifier
MR	movie review	11k	11k	1	1
CR	product review	4k	4k	1	1
SUBJ	subjectivity status	10k	10k	1	1
MPQA	opinion-polarity	11k	11k	1	1
SST	binary sentiment analysis	67k	1.8k	1	1
SST	fine-grained sentiment analysis	8.5k	2.2k	1	1
TREC	question-type classification	6k	0.5k	1	1
SICK-E	natural language inference	4.5k	4.9k	1	1
SNLI	natural language inference	550k	9.8k	1	1
MRPC	paraphrase detection	4.1k	1.7k	1	1

SentEval (Facebook researcher)

Model	MR	CR	SUBJ	MPQA	SST-2	TREC	MRPC	Avg.
Avg. GloVe embeddings	77.25	78.30	91.17	87.85	80.18	83.00	72.87	81.52
Avg. BERT embeddings	78.66	86.25	94.37	88.66	84.40	92.80	69.45	84.94
BERT CLS-vector	78.68	84.85	94.21	88.23	84.13	91.40	71.13	84.66
InferSent - GloVe	81.57	86.54	92.50	90.38	84.18	88.20	75.77	85.59
Universal Sentence Encoder	80.09	85.19	93.98	86.70	86.38	**93.20**	70.14	85.10
Sentence-BERT-base (Mean)	83.64	89.43	94.39	89.86	88.96	89.60	**76.00**	87.41
Sentence-BERT-large (Mean)	84.88	90.07	94.52	90.33	90.66	87.40	75.94	87.69
DefSent-BERT-base (CLS)	80.94	87.57	94.59	89.98	85.78	89.73	73.82	86.06
DefSent-BERT-large (CLS)	85.79	90.54	**95.58**	90.15	**91.17**	90.47	73.74	88.20
DefSent-RoBERTa-base (CLS)	83.94	90.44	94.05	90.70	89.16	90.80	75.52	87.80
DefSent-RoBERTa-large (Mean)	**86.47**	**91.53**	95.02	**91.15**	90.77	92.33	73.91	**88.74**

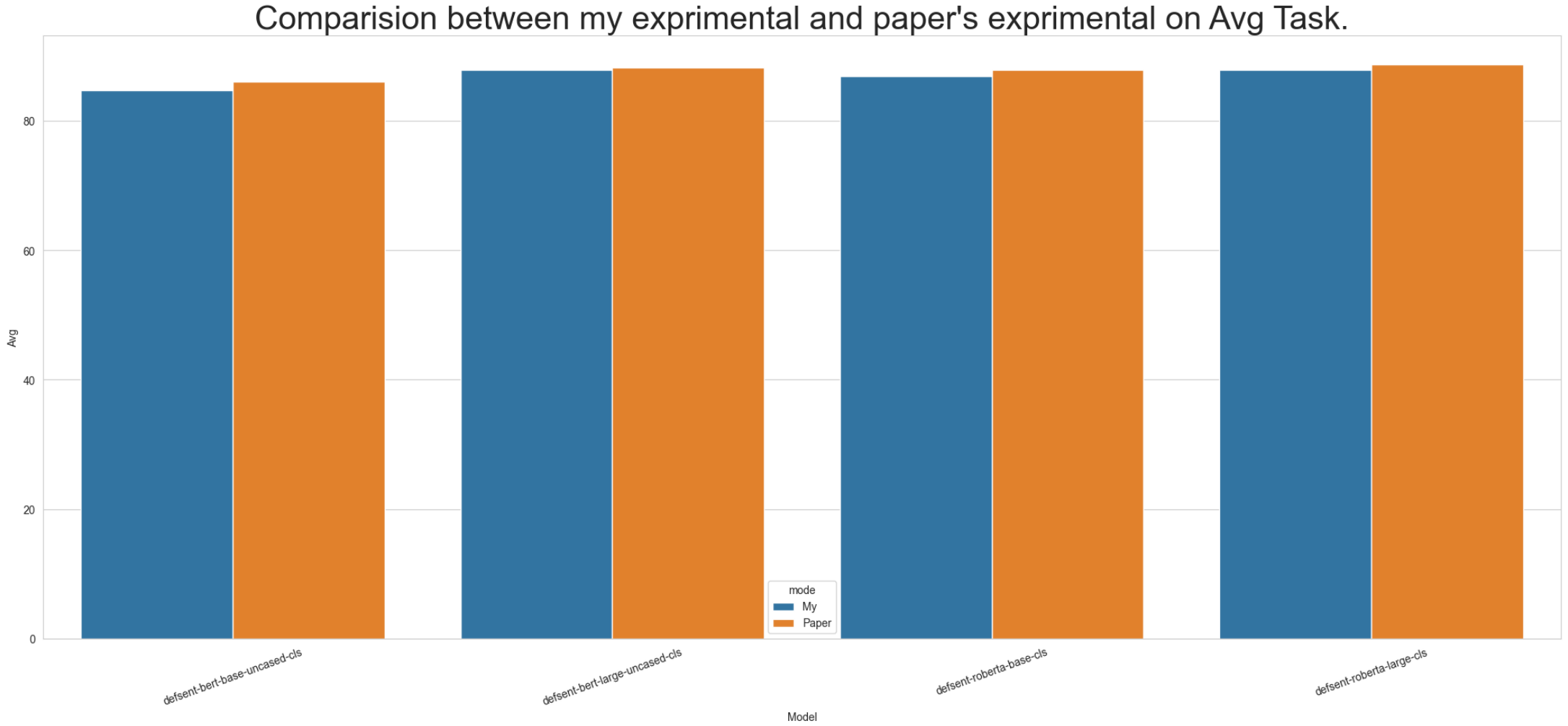
Paper experiment

SentEval (Facebook researcher)

model	MR_acc	CR_acc	SUBJ_acc	MPQA_acc	SST2_acc	TREC_acc	MRPC_acc	SICKEntailment_acc	avg_acc
defsent-bert-base-uncased-cls	80.64	86.6	93.78	89.37	85.45	84.2	72.93	81.14	84.26375
defsent-bert-base-uncased-mean	81.53	87.31	94.37	89.68	86.16	88.8	75.77	82.61	85.77875
defsent-bert-base-uncased-max	80.5	86.49	94.03	89.73	85.56	85.8	73.33	82.24	84.71
defsent-bert-large-uncased-cls	85.24	90.09	95.06	89.7	90.55	91	73.62	81.73	87.12375
defsent-bert-large-uncased-mean	84.39	88.95	94.8	89.86	89.95	87.8	74.61	78.91	86.15875
defsent-bert-large-uncased-max	83.08	88.93	93.56	89.74	87.31	84.6	76.17	81.73	85.64
defsent-roberta-base-cls	83.55	88.79	92.91	90.54	89.95	87.6	74.9	81.57	86.22625
defsent-roberta-base-mean	84.16	89.46	93.88	90.07	89.9	88.6	75.88	81.69	86.705
defsent-roberta-base-max	83.79	88.85	93.29	90.45	89.46	92	77.91	82.59	87.2925
defsent-roberta-large-cls	84.45	89.09	94.31	90.91	90.06	93.6	72.87	82.32	87.20125
defsent-roberta-large-mean	85.03	90.81	94.71	90.9	90.06	92.8	73.33	80.82	87.3075
defsent-roberta-large-max	84.36	90.38	93.84	91.04	89.13	86.8	76.75	82.57	86.85875

My experiment

SentEval (Facebook researcher)



Link(Repo)

Github

Introduction

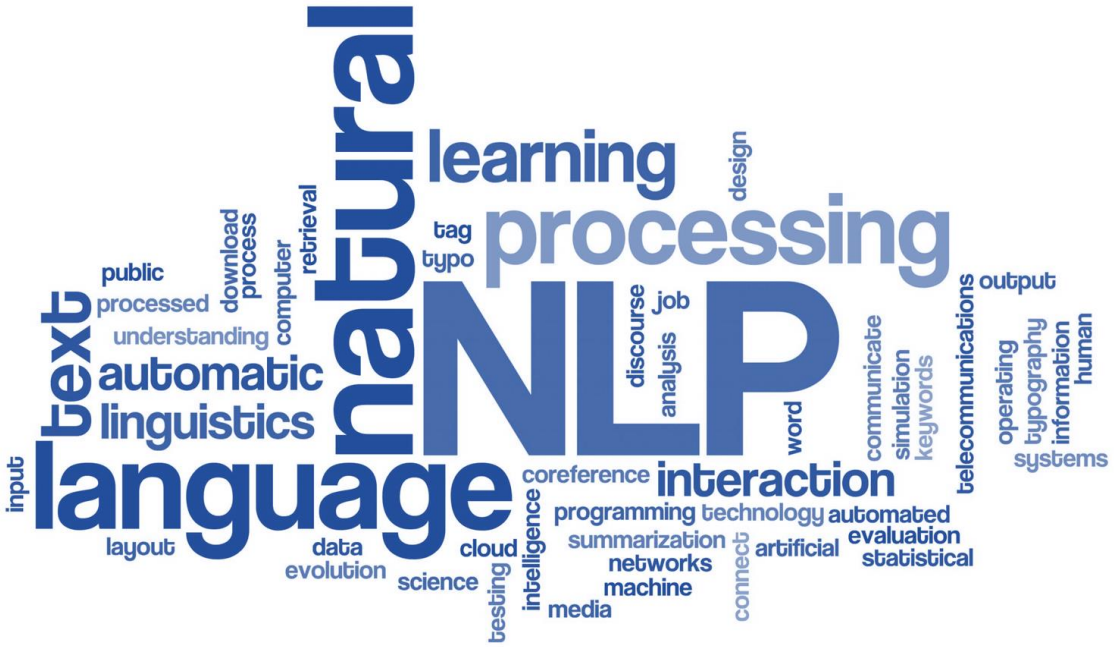
Method

Experiment

Conclusion

Conclusion

- In this paper, DefSent, a sentence embedding method using a dictionary, and demonstrated its effectiveness through a series of experiments.
- DefSent is based on dictionaries developed for many languages, so it does not require new language resources when applied to other languages.
- In future work, we will evaluate the performance of DefSent when it is applied to languages other than English.



Thanks for your listening