

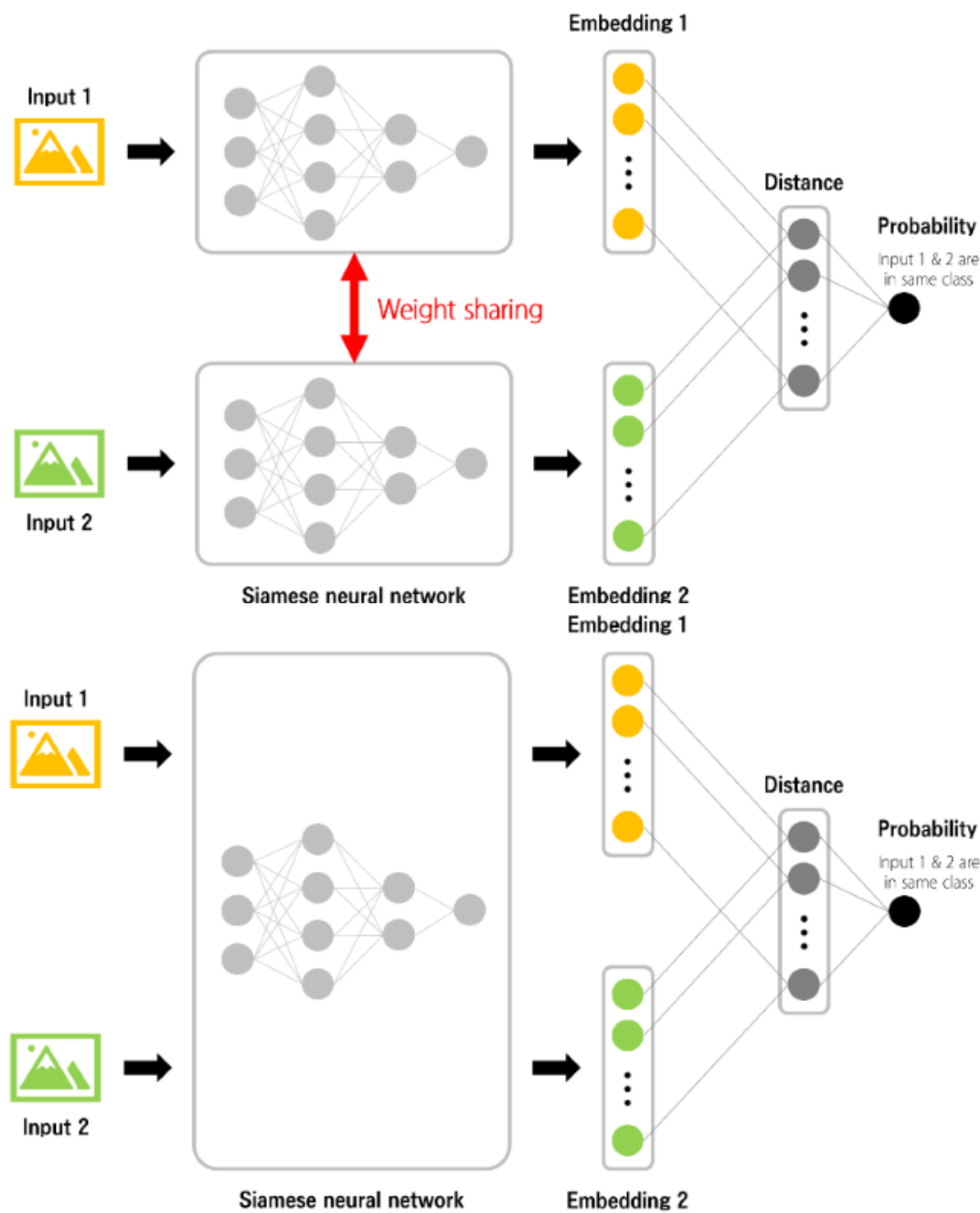
Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks

1. Introduction

- 2019, sentence embedding BERT Siamese Network
- BERT, pair regression
- Output, [CLS], GloVe
- BERT Siamese Network Sentence-BERT

2. Background

2.1 Siamese Network Architecture



- Siamese Network weight

- (Input 1, Input 2)가
- 가 Embedding 1, Embedding 2
- L1 norm, L2 norm ,
- 가 , Siamese Neural Network

3. Model

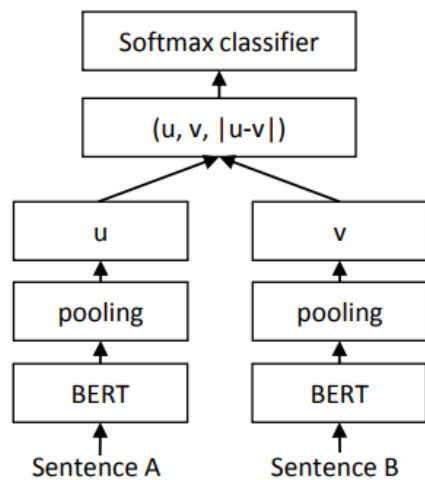


Figure 1: SBERT architecture with classification objective function, e.g., for fine-tuning on SNLI dataset. The two BERT networks have tied weights (siamese network structure).

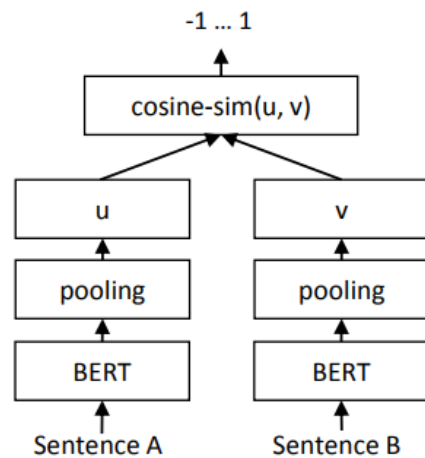


Figure 2: SBERT architecture at inference, for example, to compute similarity scores. This architecture is also used with the regression objective function.

- SBERT A B가 BERT pooling u
- SBERT v
- SBERT Classification Objective Funtion, Regression Objective Funtion, Triplet Objective Funtion

3.1 Classification Objective Funtion

Figure 1

- (1) BERT , MEAN pooling
- (2) pooling u v element wise difference $|u-v|$
- (3) u v $|u-v|$ trainable weight W_t
- (4) softmax , cross-entropy loss weight

$$o = \text{softmax}(W_t(u, v, |u - v|))$$

3.2 Regression Objective Funtion

Figure 2

- (1) Classification Objective Funtion pooling u v
- (2) u v cosine-similarity
- (3) MSE loss weight

3.3 Triplet Objective Funtion

(1) anchor sentence a positive sentence p, negative sentence n

- (2) a p 가 가 , a n 가
- (3) sx , epsilon sp 가 sn sa 가 margin
- (4) Euclidean distance , epsilon 1

$$\max(\|s_a - s_p\| - \|s_a - s_n\| + \epsilon, 0)$$

3.4 Training Details

Dataset
 SNLI : 570,000 sentence pairs
 MultiNLI : 430,000 sentence pairs

Parameter
 batch size = 16
 optimizer = Adam
 learning rate = 2e-5
 pooling strategy = MEAN

4. Evaluation - Semantic Textual Similarity

NLI (Natural Language Inference) Dataset

Example	English Translation	Label
P: 저는, 그냥 알아내려고 거기 있었어요. H: 이해하려고 노력하고 있었어요.	I was just there just trying to figure it out. I was trying to understand.	Entailment
P: 저는, 그냥 알아내려고 거기 있었어요. H: 나는 처음부터 그것을 잘 이해했다.	I was just there just trying to figure it out. I understood it well from the beginning.	Contradiction
P: 저는, 그냥 알아내려고 거기 있었어요. H: 나는 돈이 어디로 갔는지 이해하려고 했어요.	I was just there just trying to figure it out. I was trying to understand where the money went.	Neutral

STS (Semantic Textual Similarity) Dataset

Example	English Translation	Label
한 남자가 음식을 먹고 있다. 한 남자가 뭔가를 먹고 있다.	A man is eating food. A man is eating something.	4.2
한 비행기가 착륙하고 있다. 애니메이션화된 비행기 하나가 착륙하고 있다.	A plane is landing. A animated airplane is landing.	2.8
한 여성이 고기를 요리하고 있다. 한 남자가 말하고 있다.	A woman is cooking meat. A man is speaking.	0.0

- NLI Dataset (Entailment), (Contradiction), (Neutral) Classification
- STS Dataset (Similarity) 0~5 Regression
- STS Dataset SBERT 가 , spearman correlation (Unsupervised STS) (Supervised STS)

4.1 Unsupervised STS

Model	STS12	STS13	STS14	STS15	STS16	STSb	SICK-R	Avg.
Avg. GloVe embeddings	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	52.86	66.75	62.15	72.77	66.87	68.03	65.65	65.01
Universal Sentence Encoder	64.49	67.80	64.61	76.83	73.18	74.92	76.69	71.22
SBERT-NLI-base	70.97	76.53	73.19	79.09	74.30	77.03	72.91	74.89
SBERT-NLI-large	72.27	78.46	74.90	80.99	76.25	79.23	73.75	76.55
SRoBERTa-NLI-base	71.54	72.49	70.80	78.74	73.69	77.77	74.46	74.21
SRoBERTa-NLI-large	74.53	77.00	73.18	81.85	76.82	79.10	74.29	76.68

- Unsupervised STS STS 2012-2016, STS benchmark, SICK-Relatedness Dataset

- SBERT, SRoBERTa가 SICK-R task SOTA
- BERT [CLS] 가 , BERT GloVe embedding

4.2 Supervised STS

Model	Spearman
<i>Not trained for STS</i>	
Avg. GloVe embeddings	58.02
Avg. BERT embeddings	46.35
InferSent - GloVe	68.03
Universal Sentence Encoder	74.92
SBERT-NLI-base	77.03
SBERT-NLI-large	79.23
<i>Trained on STS benchmark dataset</i>	
BERT-STSB-base	84.30 ± 0.76
SBERT-STSB-base	84.67 ± 0.19
SRoBERTa-STSB-base	84.92 ± 0.34
BERT-STSB-large	85.64 ± 0.81
SBERT-STSB-large	84.45 ± 0.43
SRoBERTa-STSB-large	85.02 ± 0.76
<i>Trained on NLI data + STS benchmark data</i>	
BERT-NLI-STSB-base	88.33 ± 0.19
SBERT-NLI-STSB-base	85.35 ± 0.17
SRoBERTa-NLI-STSB-base	84.79 ± 0.38
BERT-NLI-STSB-large	88.77 ± 0.46
SBERT-NLI-STSB-large	86.10 ± 0.13
SRoBERTa-NLI-STSB-large	86.15 ± 0.35

- STS , STSB NLI 37
- STS SBERT가 가
- SBERT SRoBERTa

5. Evaluation - SentEval

Model	MR	CR	SUBJ	MPQA	SST	TREC	MRPC	Avg.
Avg. GloVe embeddings	77.25	78.30	91.17	87.85	80.18	83.0	72.87	81.52
Avg. fast-text embeddings	77.96	79.23	91.68	87.81	82.15	83.6	74.49	82.42
Avg. BERT embeddings	78.66	86.25	94.37	88.66	84.40	92.8	69.45	84.94
BERT CLS-vector	78.68	84.85	94.21	88.23	84.13	91.4	71.13	84.66
InferSent - GloVe	81.57	86.54	92.50	90.38	84.18	88.2	75.77	85.59
Universal Sentence Encoder	80.09	85.19	93.98	86.70	86.38	93.2	70.14	85.10
SBERT-NLI-base	83.64	89.43	94.39	89.86	88.96	89.6	76.00	87.41
SBERT-NLI-large	84.88	90.07	94.52	90.33	90.66	87.4	75.94	87.69

- SentEval task sentence embedding 가
- task 7 task(MR, CR, SUBJ) Logistic regression classifier , SBERT가 task

6. Ablation Study

	NLI	STSb
<i>Pooling Strategy</i>		
MEAN	80.78	87.44
MAX	79.07	69.92
CLS	79.80	86.62
<i>Concatenation</i>		
(u, v)	66.04	-
$(u - v)$	69.78	-
$(u * v)$	70.54	-
$(u - v , u * v)$	78.37	-
$(u, v, u * v)$	77.44	-
$(u, v, u - v)$	80.78	-
$(u, v, u - v , u * v)$	80.44	-

- pooling NLI STSb MEAN 가
- Concatenation u, v lu-vl 가

7. Computational Efficiency

Model	CPU	GPU
Avg. GloVe embeddings	6469	-
InferSent	137	1876
Universal Sentence Encoder	67	1318
SBERT-base	44	1378
SBERT-base - smart batching	83	2042

- Glove InferSent가 CPU , 가 , GPU SBERT가

8. Conclusion

- Network BERT sentence embedding , Simaese
- SBERT SBERT sentence embedding ,
- BERT

clipboard-202208021603-s80w4.png	63 KB	2022-08-02
clipboard-202208021604-vqmy3.png	50.2 KB	2022-08-02
clipboard-202208031258-czhro.png	5.96 KB	2022-08-03
clipboard-202208031259-wuqqg.png	5.17 KB	2022-08-03
clipboard-202208031324-v7p3s.png	72.9 KB	2022-08-03
clipboard-202208051029-xqlop.png	48.1 KB	2022-08-05
clipboard-202208051030-1n5ux.png	22.1 KB	2022-08-05
clipboard-202208051526-bmznr.png	104 KB	2022-08-05
clipboard-202208051527-qxmrx.png	106 KB	2022-08-05
clipboard-202208051716-cszrl.png	105 KB	2022-08-05
clipboard-202208051719-5npfl.png	98.6 KB	2022-08-05
clipboard-202208051723-8txyl.png	74.6 KB	2022-08-05