

Understanding Sentence Similarity

Arvin-Sachin Group

Arvin Goyal - 2019900039

Sachin Raja - 2020801001

Problem Statement

- Do pair of sentences convey the same meaning?
- Also known as paraphrase detection
- Output in the form of a score between 0-5, 0 indicates no similarity and 5 indicates high semantic similarity.

Why is this problem important?

- Text Summarization
- Document Clustering
- Question Answering Systems
- Natural Language Inference
- Information Retrieval
- Plagiarism Detection and so on
- Basically, its precursor to almost all NLP problems as it requires semantic understanding of the sentence.

Why is this problem hard?

- Limited Availability of annotated data
- Variable sentences length
- Complex sentence structures

Formal Problem Definition

Given a pair of sentences, X (x_1, x_2, \dots, x_{T_a}) and Y (y_1, y_2, \dots, y_{T_b}) where both sentences have different lengths, denoted by T_a and T_b , the goal is to predict a similarity score between these two sentences.

This problem is generally tackled as a supervised regression or classification problem.

Related Work

Word-Based Similarity

- Max Similarity
- Similarity Matrix
- Analysis of similar and dissimilar parts of sentences
- Word sense disambiguation

Structure-Based Similarity

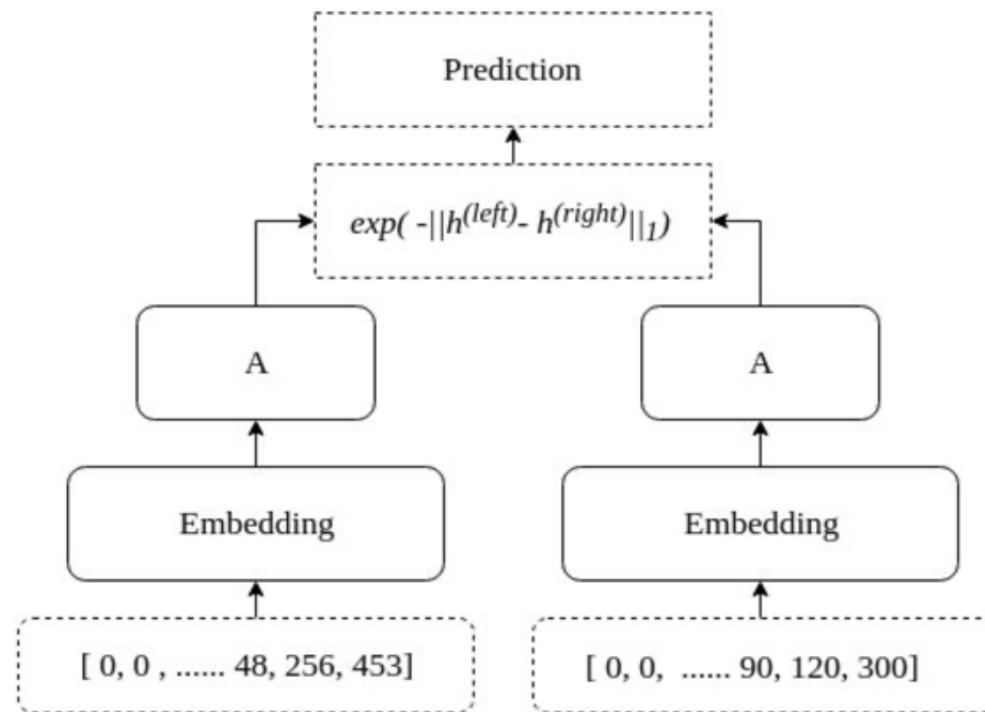
- Grammar Based
- Using parts of speech
- Using word order

Vector-Based Similarity

- Distributional Sentence Similarity
- Average of word vectors
- Learning Based Vectors

Methodology:

Base line Architecture



Distance Measures

- Euclidean Distance

$$\exp^{-\left\| \frac{s_1}{\|s_1\|_2} - \frac{s_2}{\|s_2\|_2} \right\|_2}$$

- Manhattan Distance

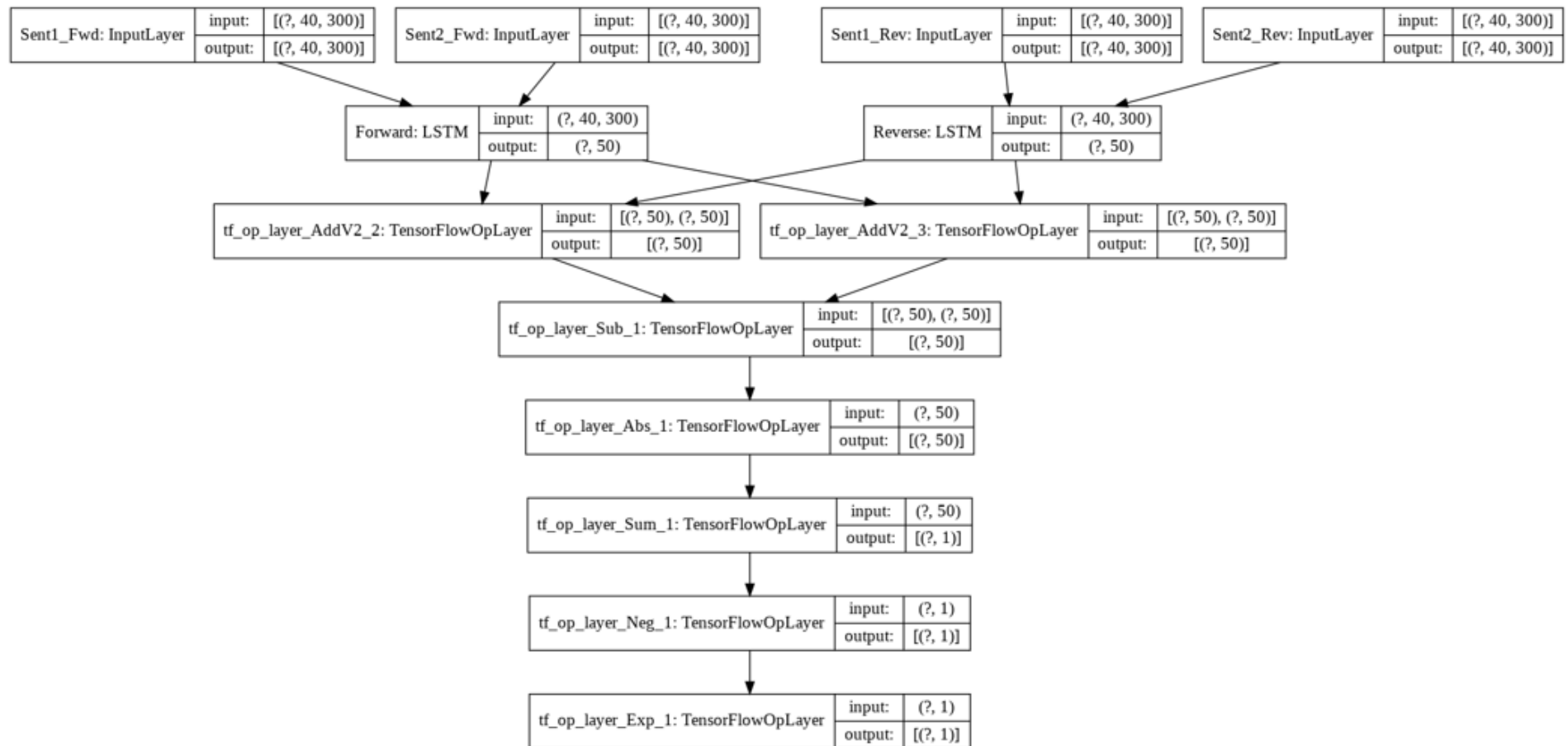
$$\exp^{-ABS\left\| \frac{s_1}{\|s_1\|_2} - \frac{s_2}{\|s_2\|_2} \right\|_1}$$

- Cosine Distance

$$1 - \left(\frac{s_1}{\|s_1\|_2} * \frac{s_2}{\|s_2\|_2} \right)$$

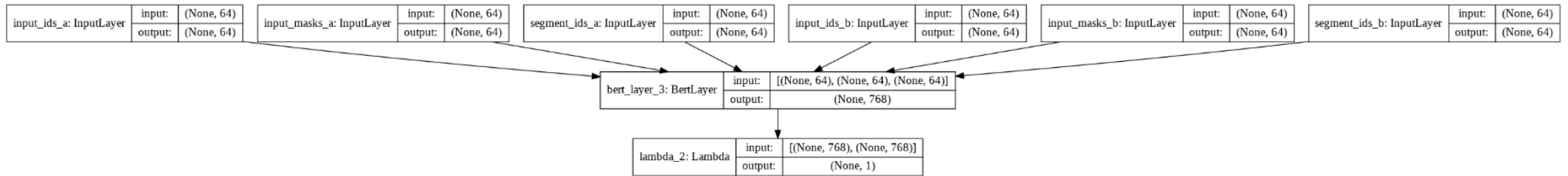
Methodology:

Extension: Bi-directional LSTM



Methodology:

Extension: Self and cross attention using BERT



RESULTS AND ANALYSIS:

MSRPC DATASET

Classification report				
	Precision	Recall	F1-score	Support
0	0.65	0.17	0.26	578
1	0.69	0.95	0.8	1147
Accuracy			0.69	1725
Macro avg	0.67	0.56	0.53	1725
Weighted avg	0.68	0.69	0.62	1725

Confusion Matrix		
	Actual 1	Actual 0
Predicted 1	96	482
Predicted 0	52	1095

RESULTS AND ANALYSIS:

SICK DATASET

Model Detail	Pearson Score	Spearman Corr	MSE	1 epoch training time	1500 sample inference time
Manhattan Distance using Unidirectional LSTM (base model)	0.501	0.474	1.278	5.0 sec	0.525 sec
Manhattan Distance + Bidirectional LSTM (base model)	0.826	0.759	0.351	3.0 sec	0.458 sec
Manhattan Distance + Bidirectional LSTM + l2 normalization	0.852	0.795	0.292	3.0 sec	0.505 sec
Euclidean Distance + Bidirectional LSTM + l2 normalization	0.863	0.809	0.268	3.0 sec	0.458 sec
Cosine similarity + Bidirectional LSTM + l2 normalization	0.855	0.802	0.284	2.0 sec	0.436 sec
Two parallel SBERT Embedding(one for each sentence) + Manhattan Distance + norm	0.779	0.695	0.4	912 sec	79.25 sec
Single BERT (cross sentence attention) + Manhattan Distance on output sequence	0.874	0.829	0.253	741 sec	53.09 sec
Single BERT (cross sentence attention) + pooled output + norm + <u>relu</u> + <u>exp</u>	0.887	0.835	0.214	711 sec	57.95 sec
Single BERT (cross sentence attention) + pooled output + norm + sigmoid	0.898	0.848	0.203	790 sec	63.97 sec

QUALITATIVE ANALYSIS ON SICK DATASET:

COMPARISON OF DISTANCE MEASURES

Sentence A	Sentence B	GT	L1 Sim	L2 Sim	Cos Sim	Deviation	
			Pred	Pred	Pred	Inter-pred	GT
A child is playing with a water spout outdoors and the rest of his family is watching	A man is standing beside a birdcage which is large and colorful	1.40	3.42	1.53	1.49	0.90	1.17
Three small dogs are sniffing at something	Butter is being chopped into a container by a man	1.00	1.44	1.79	3.11	0.72	1.33
A man is shooting a shotgun	Someone is playing the guitar	1.00	1.57	1.40	2.98	0.71	1.21
The sun is shining on the face of the girl who is in a yellow dress	A young black child is standing on the edge of a body of water near some buckets	2.00	2.34	3.48	1.81	0.70	0.88
A person is rinsing a steak with water	A man is buttering a slice of bread	1.70	2.95	1.26	2.31	0.69	0.84
A woman is taking two eggs out of a pot of water	A person is buttering a tray	1.80	3.19	1.57	2.30	0.66	0.86
A bike rider in a black and red uniform is standing on a dirt bike	Five adults are sitting on stone steps	1.10	3.09	1.48	2.24	0.66	1.34
A homeless man is holding up a sign and is begging for money	The cat is hungrily drinking milk	1.00	2.12	0.65	1.74	0.62	0.80
A man with tattoos is lounging on a couch and is looking for a pencil	A cow is eating hay	1.00	2.46	1.09	2.26	0.61	1.12
One boy in orange shorts is standing on a rock cliff over the water and the other boy in black shorts is jumping of it into the water	A baby is sneezing and frightening another baby	1.00	2.66	1.22	1.81	0.59	1.07

QUALITATIVE ANALYSIS ON SICK DATASET:

COMPARISONS OF BERT IMPLEMENTATIONS

Sentence A	Sentence B	GT	SBERT Pred	BERT Seq L1 dist Pred	BERT Pool Sig Pred	Deviation	
						Inter-pred	GT
Two crocodiles are floating in a green colored swimming pool near some playing kids	Two poodles are in the snow and one is jumping high	1.10	3.56	1.80	1.36	0.95	1.48
The brown dog is playing outdoors	Two people are arguing near a crowd	1.00	3.58	1.91	1.36	0.94	1.59
Two dogs are playing on the beach	Two people are driving a jeep and a lady is sitting on the top of it	1.00	3.52	1.88	1.37	0.92	1.56
Two men are walking through the grass	Two men are standing in deep water	2.40	3.82	2.12	1.71	0.91	0.93
The people are walking on the road beside a beautiful waterfall	There is no brown dog and black dog playing in the sand	1.00	3.52	1.81	1.42	0.91	1.55
A large brown dog is jumping over a red hurdle	A young man is running away from the fishing line	1.00	3.39	1.81	1.37	0.87	1.47
A person is kicking a soccer ball between their feet	One brown and black dog is running through the leaves	1.00	3.37	1.81	1.36	0.86	1.46
There is no girl playing a flute	Two baby pandas are playing	1.00	3.39	1.90	1.37	0.86	1.49
A woman is cutting a fish	The man is slicing potatoes	2.00	3.38	1.88	1.37	0.85	0.88
A boy is standing in a room by a lamp light	A small group of people are standing and two are sitting on the couch	2.20	3.59	1.84	1.73	0.85	0.87

QUALITATIVE ANALYSIS ON SICK DATASET:

COMPARISON OF BI_DIRECTIONAL LSTM AND BERT WITH CROSS-ATTENTION

Sentence A	Sentence B	GT	Manh Pred	BERT Pool Sig Pred	Deviation	
					Inter-pred	GT
Different things from a war are being shown to some people by a veteran	Three friends are making faces for the camera	1.40	3.51	1.38	1.06	1.49
A person is scrubbing a zucchini	The woman is cutting cooked octopus	1.19	3.30	1.39	0.95	1.50
A child is playing with a water spout outdoors and the rest of his family is watching	A man is standing beside a birdcage which is large and colorful	1.40	3.42	1.56	0.93	1.43
A boy is standing next to the opening of a fountain	The man is standing on a rocky mountain and gray clouds are in the background	1.40	3.25	1.45	0.90	1.31
Two crocodiles are floating in a green colored swimming pool near some playing kids	Two poodles are in the snow and one is jumping high	1.10	3.11	1.36	0.88	1.44
A man is skateboarding on a half pipe	The kids are playing outdoors near a man with a smile	1.60	3.28	1.59	0.85	1.19
A bike rider in a black and red uniform is standing on a dirt bike	Five adults are sitting on stone steps	1.10	3.09	1.40	0.85	1.42
A woman is eating a cupcake	A man is slicing a potato	1.30	3.04	1.37	0.83	1.23
A group of children in a church basement is playing maracas and tambourines	Tambourines are being played by a group of children	4.60	2.84	4.46	0.81	1.25
A horse is standing	Someone is holding a hedgehog	1.00	2.97	1.37	0.80	1.42

Observations

- Correlation scores is the best evaluation metric since instead of absolute value matching, we are more interested in the relative ranking of sentence pairs.
- Normalizing sentence vectors before similarity calculation improved correlation scores by an average of 3 percent.
- BERT with self and cross attentions performed significantly better on longer sequences.
- While Siamese network for sentence similarity paper reports best results using Manhattan distance, we obtained best results with Euclidean distance after normalizing the sentence vectors.
- Both BERT and LSTM based models resulted in more false positives than false negatives.

Thank you!

- Questions please