Natural Language and Processing

Siamese Networks

Utkarsh Agiwal

Summer of Science, July 2022

What are they?

Often while asking questions on websites like Quora, Stackoverflow, etc. you tend to get substitutes for your question, or questions similar to yours. This is done with the help of siamese networks. They identify similarity between things. For example, 'What is your age' and 'How old are you?' actually are question duplicates.
Applications:

- Measure similarity between two classes
- Helpful in Image Classification tasks

Model Architecture

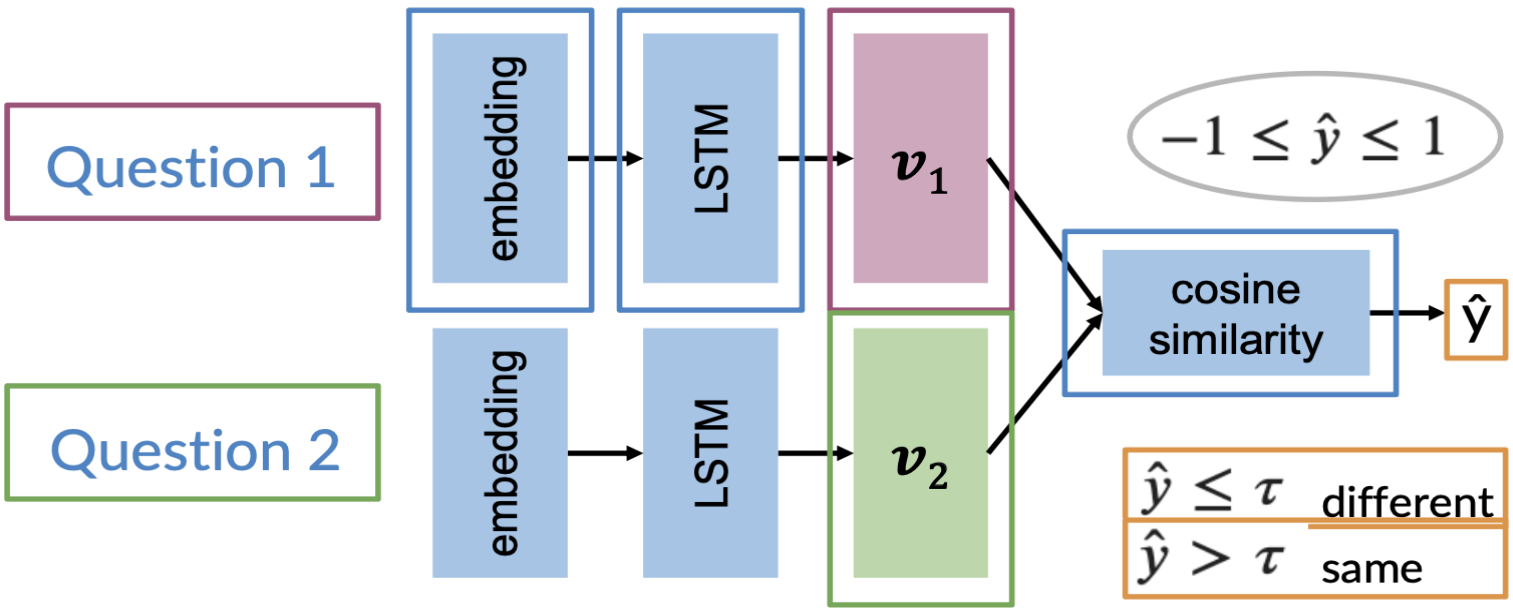


Figure: Siamese Architecture

In Siamese network the basic network for getting features of entities(images/text) is kept same and the two entities we want to compare are passed through the exact same network. By the exact same network it is meant that both the entities are passed through the same architecture having same weights as shown in the figure. At the end of common network a vectored representation of the input is obtained which can then be used for measuring or quantifying the similarity between them.

Loss Function

An anchor statement is chosen which is used to designate similarity to sentences.

- Anchor : How old are you?
- Positive : What is your age?
- Negative : Where are you from?

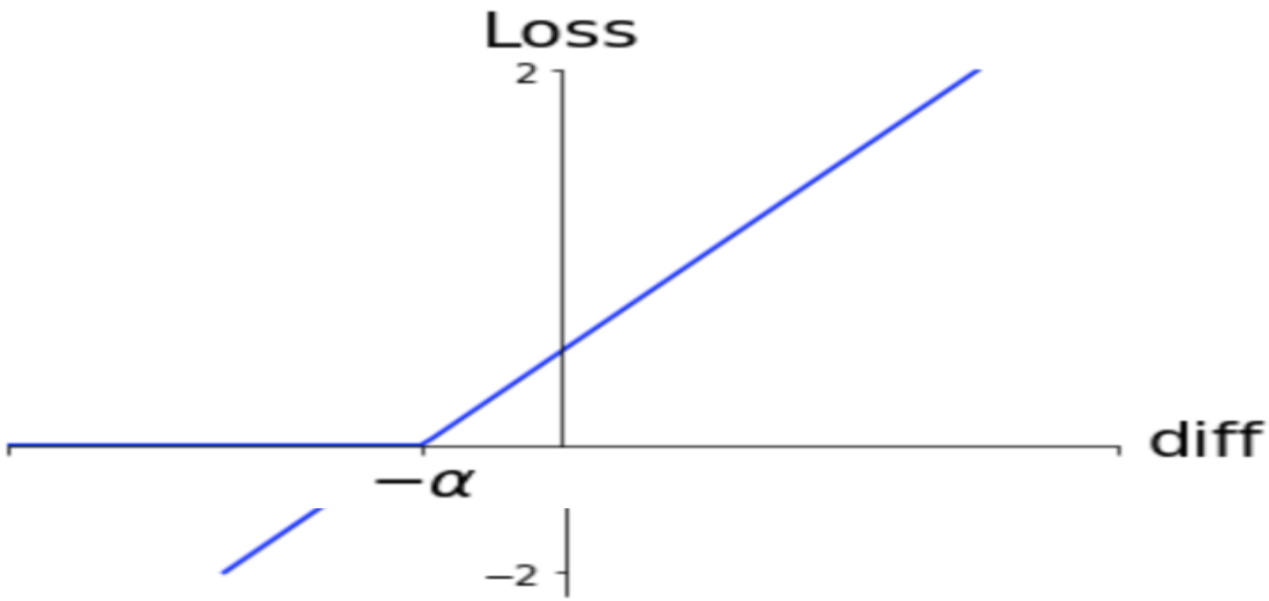
Similarity between anchor and positive statements $s(A, P)$ as well as between anchor and negative statements $s(A, N)$ is calculated using the cosine similarity function. Loss function is defined as:

$$diff = s(A, N) - s(A, P) \tag{1}$$

Triplet Loss

$$L = \begin{cases} 0 \text{ if diff} + \alpha \leq 0 \\ \text{diff if diff} + \alpha > 0 \end{cases} \tag{2}$$

The triplet loss adds a margin since we want our model to train more over examples which have diff close to zero, since these are examples which are highly correlated and might be important.



Cost Function and Testing

$$J = \sum_{i=1}^{m} L(A^{(i)}, P^{(i)}, N^{(i)}) \tag{3}$$

Testing:

- Convert each input into an array of numbers
- Feed arrays into your model
- Compare v_i, v_2 using cosine similarity
- Test again a threshold