Features Used: -

ROSE features as given in “Regression and Ranking based Optimisation for Sentence Level Machine Translation Evaluation by Xingyi Song et al.” Below are the features used.



I have formulated this problem as a classification problem instead of a pairwise ranking problem, which seems to be the first intuition. **The input to the classification system is the vector x2 - x1, where**

**x**1 :- ROSE feature vector calculated for sentence 1 and reference sentence.

**x**2 :- ROSE feature vector calculated for sentence 2 and reference sentence.

**As proved in ([Herbrich 1999](http://www.mendeley.com/research/support-vector-learning-ordinal-regression/)), the pairwise ranking problem can be transformed into a two-class classification problem. Thus, I am using the difference vector as input to the classification model.**

Rather than converting it into a two-class classification problem, I have converted it into a 3-class classification problem. For classification, I am using a 3-layer FC network with a softmax layer after the last FC layer. For non linearity, ReLu is used after first two FC layers.

The input to the Neural Net is a vector of size 33.

First Layer :- Weight vector of size 33x16

Second layer :- Weight vector of size 16x8

Third layer :- Weight vector of size 10x3(2 semantic features are input directly to last FC layer)

The features input to the first layer are syntactic in nature. To improve the classification score we need semantic features too. Thus, we use skip features in the last FC layer. I am using spcay package’s (NLP) Doc similarity function to get the similarity score between the sentences and reference sentence. Spacy package uses inbuilt word vector embeddings to calculate this score. This feature is used as a semantic feature to help classifying the sentences in a better way.

This Neural Network design is inspired from following paper:- “Machine Translation Evaluation with Neural Networks , Guzman et al.” <https://arxiv.org/pdf/1710.02095.pdf>

The Neural Net generates three outputs which is then converted into probabilities using the softmax function and thus used to classify the sentence as -1,0,1.