In task two we have to answer the question from a retrieved paragraph in task 1. It is a span based machine reading comprehension task were model will take a passage and question <P,Q> and predict the start index and end index of the answer in the given paragraph. The architectural diagram of the proposed system is given in the figure. The system can predict start index and end index if the question is answerable else it will return a null string. The briefe description of each module of the system given below.

Encoder

Pre-trained Language Models(PrLM) like BERT or ALBERT is used as an encoder. The passage and question are first concatenated and tokenized (used pre-trained tokenizers corresponding to pre-trained language model used) to generate input embedding X={x\_1,x\_2….x\_n}. So the input embedding will be the sum of its token embedding, position embedding, and token-type embedding. The input embeddings are then fed in to the encoder to generate contextual representation. Here hi is a vector representing each token having a size equal to the number of hidden units in the last layer of the pretrained language model.

Decoder:

Used a fully connected layer with two output units and softmax activation as the decoder. The decoder will take H(nxk) as input and predicted(output pf size - nx2) start and end probabilities s and e. Cross entropy is used as the training objective for this answer span prediction.

Internal Verifier

This module incorporates an answerability check to the training objective. A fully connected layer with two output units and softmax activation function used in this module and acts like a classifier. The pooled first token ([CLS]) representation h1 ∈ H, as the overall representation of the sequence fed into the internal verifier to predict answerable and unanswarable probabilities. Cross entropy is used as the training objective of this module.

Answerability Verification

We used threshold-based answerability verification for this task. We calculate the has-answer score (score\_has) and the no-answer score (score\_null) as follows,

If score\_diff is more than answerable threshold (delta) the model will predicts the answer span and null string otherwise. The value of (delta) is determined with the help of development set.

Experiments and Results:

We used pre-trained language models BERT and ALBERT and used pretrained weights in the encoder. Then fintuned the model with all the official hyperparameters of the pretrained language models. Initially we finetuned the model without internal verifier using the span loss. The start index and end index for non answerable questions used as 0 (first token).

Then model is finetuned with internal verifier with overall all loss function as the combination of both span loss and verification loss as shown below. The value of alpha\_1 and alpha\_2 are set as 1 and 0.5. The training and validation F1 scores obtained for different training methods are reported in the table.

For the training we used ADAM optimizer with initial learning rate as 2e-5 and L2 weight decay of 0.01. The batch size is used as 8 and maximum number of epochs set as 5 in all the experiments.