**Working**

Question are represented as sparse matrix i.e at all position except at the index of current token there will be 0 . The index are made using the vocabulary of all tokens present in sentences.

So Each question is represented as a matrix with V columns and Lmax Rows

V = Vocabulary i.e No of words in all sentences combined

Lmax = maximum number of tokens present in a sentence

**Embedding**

The dimensionality of the above formed matrix is reduced to De Columns and Lmax rows by multiplying it with another matrix of dimension V rows and De columns also known as Word Embedded matrix (WE) matrix .

(X embedded) Lmax X De = (X oneHot) Lmax X V \* ( WE ) V X De

**CNN Architecture**

To keep the freedom in architecture the filters are kept in (Wconv) Fs X De

Fs = Size of filter { 1, 10 }

De = Degree pf freedom { 10 , 500}

The result of convolution is X conv = X emb (Dot Product) Wconv

The Final result after activation is applied

Xactivation = F( Xconv + b conv)

where F () = activation function

b = bias

**Pooling**

The pooling layer is implemented using one max pooling where pooling layer is constituted by a horizontal vector p with dimension Ntot, whose elements pi , with i = 1,..., Ntot , are:

P i = max ( X actual) i

**Finally at Fully Connected layer**

K = number of classes required in the current case it was 9

Y = P .Wfc + bfc

P= pooling matrix

Wfc = Weights

bfc = Bias

the final output is made of only one Node i.e

output = max (Y)

**Model Optimization**

10 fold cross validation is used i.e The data is divided into 10 separate regions and each region is used separately to train and test the model .

The accuracy is calculated by averaging the output which comes in each test .