1.Suppose that we have 20 triples{T=(h,r,t)} in the **training sets** ,means we might we having at max 20 head(h) and 20 tail(t) and at max 10 relation(r) or less than that in every case.

EX.-

|  |  |  |
| --- | --- | --- |
| Management | involves | treatment |
| The World Health Organization | declared | outbreak |
| common symptoms | include | cough |

Here we have 3 heads ,3 tails and 3 relations.

2.And we have 5 triples {T=(h,r,t)} in the **test set** and the test set include all the entities and relation present in training set .But the triple is totally different.

Ex.-

Management, include, treatments

Here Management(head) and treatment (tail) and include(relation) all the three are present in training set but it is totally different triple.

3.So now we apply TransE model on the following train and test set. We set hyperparameters in the models such as :-

* **k** (int) – Embedding space dimensionality. [ here we have taken 10 ]
* **eta** (int) – The number of negatives that must be generated at runtime during training for each positive.[Suppose here we take value 5 then the scores of one true triple will be generated on the basis of 5 negative triples] .Here the triple with label 1 is true rest are false.

**h r t label**

Management incolves treatment [1]

Managemnt involves outbreak [0]

Management involves cough [0]

Management involves common symotoms [0]

Management involves The world Health organization [0]

Now the embedding model creates vectors of the triples and gives them to a scoring function ,In this case it is :-

fTransE=−||esub+epred−eobj||n

Here embedding of the Management entity will be same in all the positive and negative triples

So its gives out the scores of the triples

**h r t label scores rank**

Management incolves treatment [1] 0.753 1 \*

Managemnt involves outbreak [0] 0.789 2

Management involves cough [0] 0.695 3

Management involves common symotoms [0] 0.456 4

Management involves The world Health organization [0] 0.234 5

Scores of negative triple may be good but the model knows the label we given to that true triple which will fix rank of that triple irrelevant to the score but it depends on the model how good scoring function is to give the best accuracy.

Same goes for every following training triple in the train set.

Once training is Done for the triple we fit the model and Now we Evaluate the test set based on the Ranks of the test set .

Ex: Management, include, treatments

It will do same strategy for predicting the rank of the test triple by first corrupting the head and tail .For each test triplet, the head is removed and replaced by each of the entities in the filters we define which is generally combination of training and test set. After it generates score ,the rank is generated here as we have not given the label because we are checking the model to predict the rank of true triple based on the training set.

This is how the model is evaluated .