PROJECT

Submitted By : Rohini.N

Topic : Large Scale Hierarchical Text Classification Using Deep Graph CNN

Guide : Ms. Sunitha C

About Project :

Text classification is the task of classifying unlabelled natural language documents into a predefined set of categories. In large scale text classification, no.of classes can run into thousands and in some cases, each document may only belong to single class while in others to more than one. Considering the hierarchical relations offers extra information to a classification system which can improve scalability and accuracy. Task of classification can depend on various factors like structure of data,size of data processed etc. Many real world problems however need to consider a huge amount of data to be classified from many sources.

The proposed method uses a Graph-CNN based deep learning model to first convert texts to graph of words and then uses graph convolution operations to convolve the word graph. To leverage hierarchy of labels, a regularized deep architecture with dependency among labels is also considered. The dataset used for the system are RCV1 and NYTimes.

To convert a document to a graph, a natural way is to use the word co-occurrence. The combination of distant n-grams mined from word co-occurrence graph can compose sub-graph.We sort all nodes in a graph based on their degrees (number of neighborhoods of a node). The first convolutional layer takes the feature space of size $N \times g \times D$ as input, where N number of selected and normalized sub-graphs, g is the size of receptive field (size of the sub-graphs), and D is the dimension of word embeddings.

Publications :

Paper : Methods of Large Scale Text Classifications
Conference : ICICT 2018
Journal : IEEE
Status : Presented

Paper : Prognosis Of Hear Disease Using Data Mining Techniques
Conference : ICICT 2018
Journal : IEEE
Status : Presented