MAGNET: Multi-Label Text Classification using Attention-based Graph Neural Network

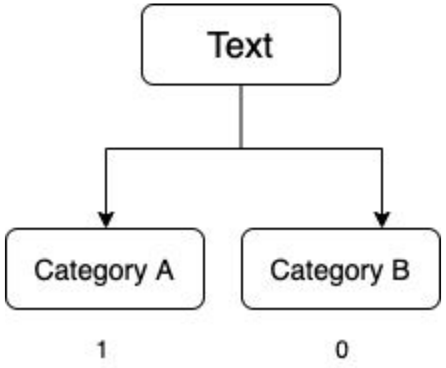
Ankit Pal
Research Engineer, Saama Tech.
ankit.pal@saama.com

Saama Technologies

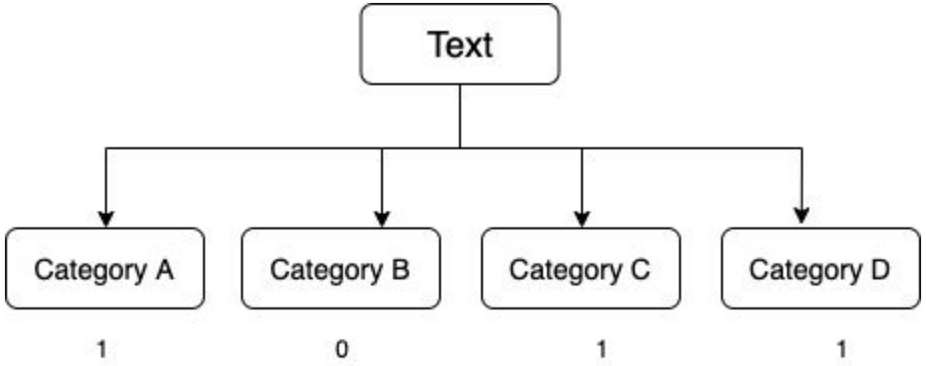
- Accelerated Drug Development
- Critical Data Insights
- Advanced Analytics

Multi-Label Text Classification (MLTC)

- Binary Text Classification

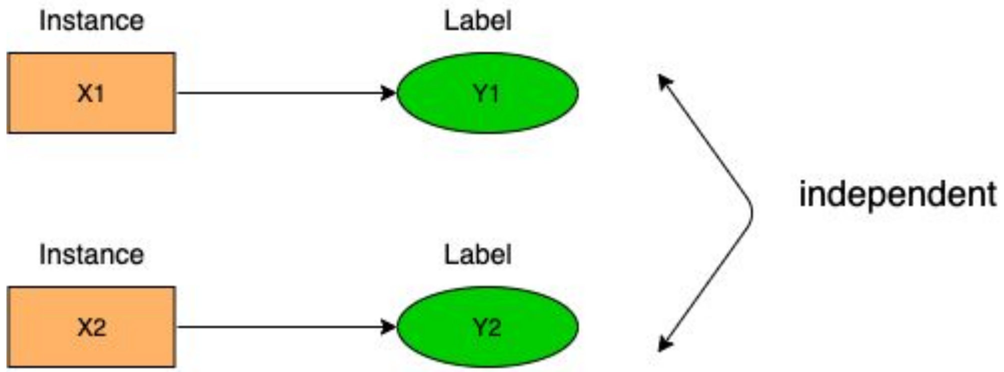


- Multi-label Text Classification

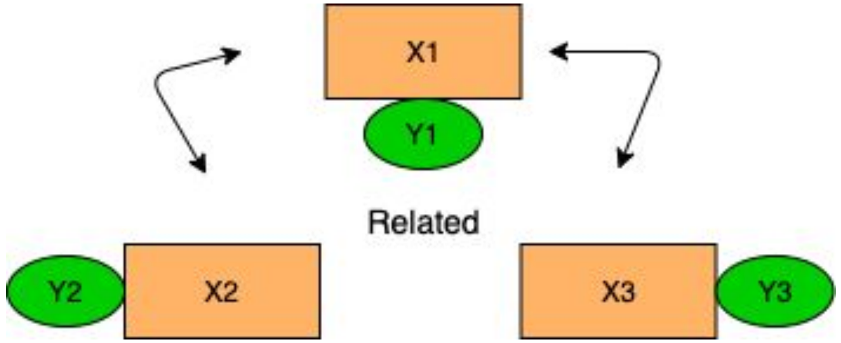


Multi-label Text Classification (MLTC)

Conventional classification approaches assume that instances are independent identically distributed



In relational data and information networks, instances are correlated with each other.



Reuters dataset

In reuters dataset, there is correlation between the labels.

Existing methods tend to ignore the relationship among labels.

What is Graph Neural Network?

Neural Network for Graph datasets :

- Understanding the connection and capturing better correlation

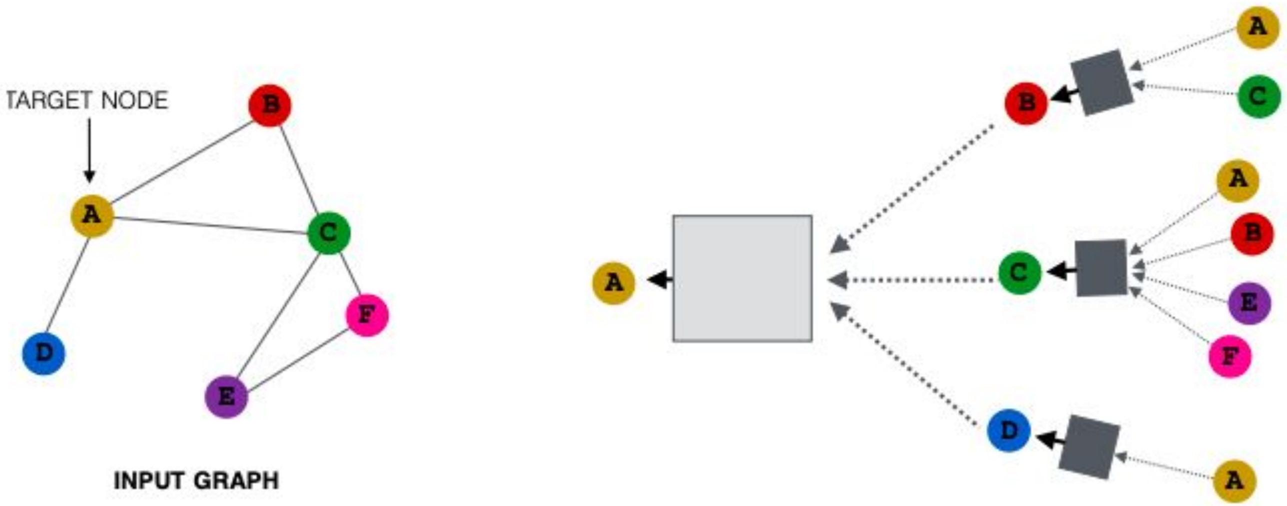
Assume we have a graph G:

- V is the vertex set.
- A is adjacency matrix (assume binary)
- X is the matrix of node features

What is Graph Neural Network?

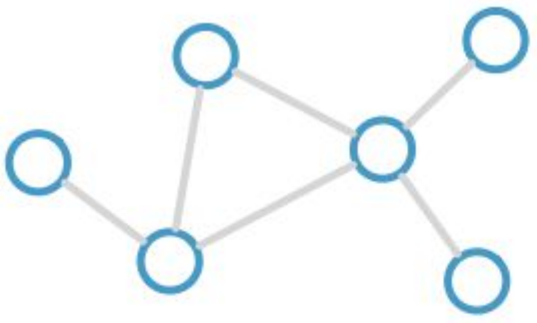
Key Idea : Neighborhood Aggregation

Generate node embeddings based on local neighborhoods.

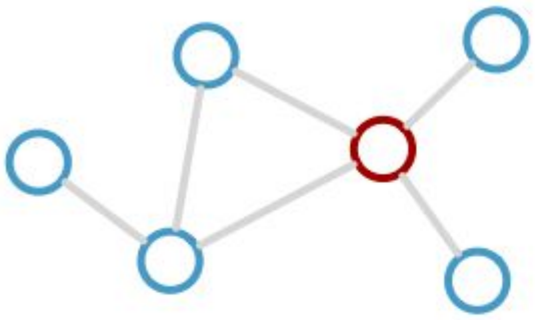


What is Graph Neural Network?

Consider this graph

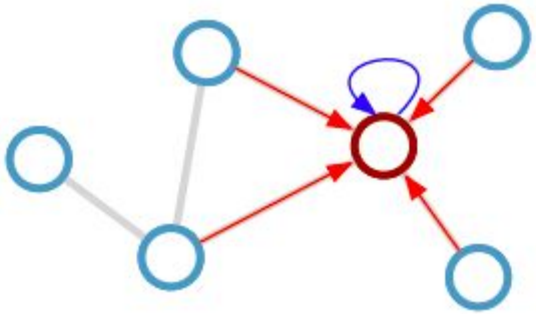


Calculate Update for red node



What is Graph Neural Network?

Calculate Update for red node

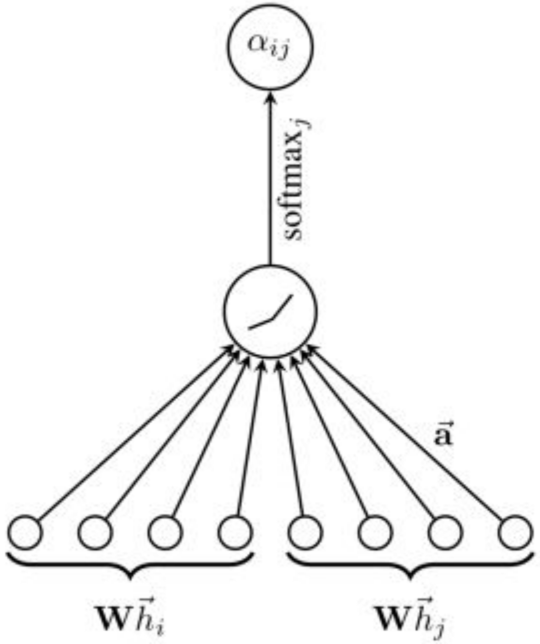


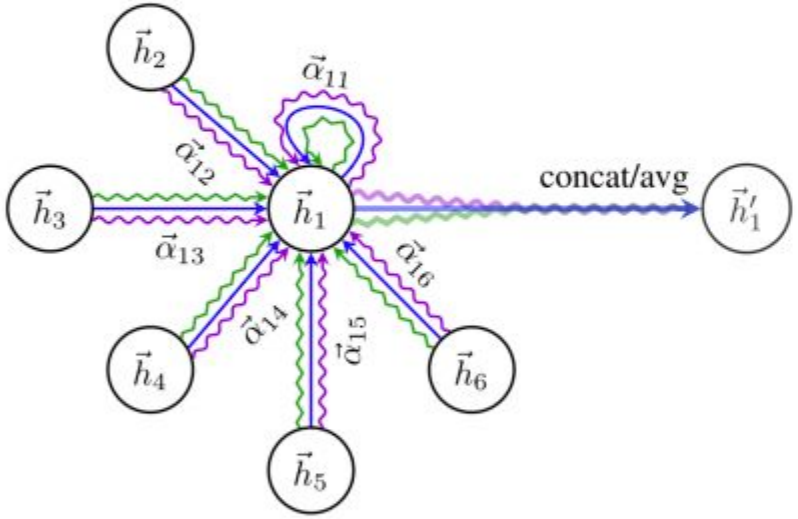
Update rule:

$$\mathbf{h}_i^{(l+1)} = \sigma\left(\mathbf{h}_i^{(l)}\mathbf{W}_0^{(l)} + \sum_{j\in\mathcal{N}_i}\frac{1}{c_{ij}}\mathbf{h}_j^{(l)}\mathbf{W}_1^{(l)}\right)$$

Graph Attention Networks

Rather than node embedding with equal or pre-defined weights of nodes, Graph attention networks utilize neighbours representations according to their relevance.





$$\vec{h}'_i = \sigma\left(\frac{1}{K}\sum_{k=1}^{K}\sum_{j\in\mathcal{N}_i}\alpha_{ij}^{k}\mathbf{W}^{k}\vec{h}_j\right)$$

$$\alpha_{ij} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i\|\mathbf{W}\vec{h}_j]\right)\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(\text{LeakyReLU}\left(\vec{\mathbf{a}}^T[\mathbf{W}\vec{h}_i\|\mathbf{W}\vec{h}_k]\right)\right)}$$

Graph Attention Networks For Multi-label Classification

MAGNET: Multi-Label Text Classification using Attention-based Graph Neural Network

Ankit Pal, Muru Selvakumar and Malaikannan Sankarasubbu
Saama AI Research, Chennai, India
{*ankit.pal, selvakumar.murugan, malaikannan.sankarasubbu*}@*saama.com*

Keywords: Multi-label Text Classification, Graph Neural Networks, Attention Networks, Deep Learning, Natural Language Processing, Supervised Learning.

Abstract: In Multi-Label Text Classification (MLTC), one sample can belong to more than one class. It is observed that most MLTC tasks, there are dependencies or correlations among labels. Existing methods tend to ignore the relationship among labels. In this paper, a graph attention network-based model is proposed to capture the attentive dependency structure among the labels. The graph attention network uses a feature matrix and a correlation matrix to capture and explore the crucial dependencies between the labels and generate classifiers for the task. The generated classifiers are applied to sentence feature vectors obtained from the text feature extraction network(BiLSTM) to enable end-to-end training. Attention allows the system to assign different weights to neighbor nodes per label, thus allowing it to learn the dependencies among labels implicitly. The results of the proposed model are validated on five real-world MLTC datasets. The proposed model achieves similar or better performance compared to the previous state-of-the-art models.

Graph Attention Networks For Multi-label Classification

MAGNET network

We present graph attention network-based model to capture the attentive dependency structure among the labels. The graph attention network uses a feature matrix and a correlation matrix to capture and explore the crucial dependencies between the labels and generate classifiers for the task. The generated classifiers are applied to sentence feature vectors obtained from the text feature extraction network(BiLSTM) to enable end-to-end training.

Graph Attention Networks For Multi-label Classification

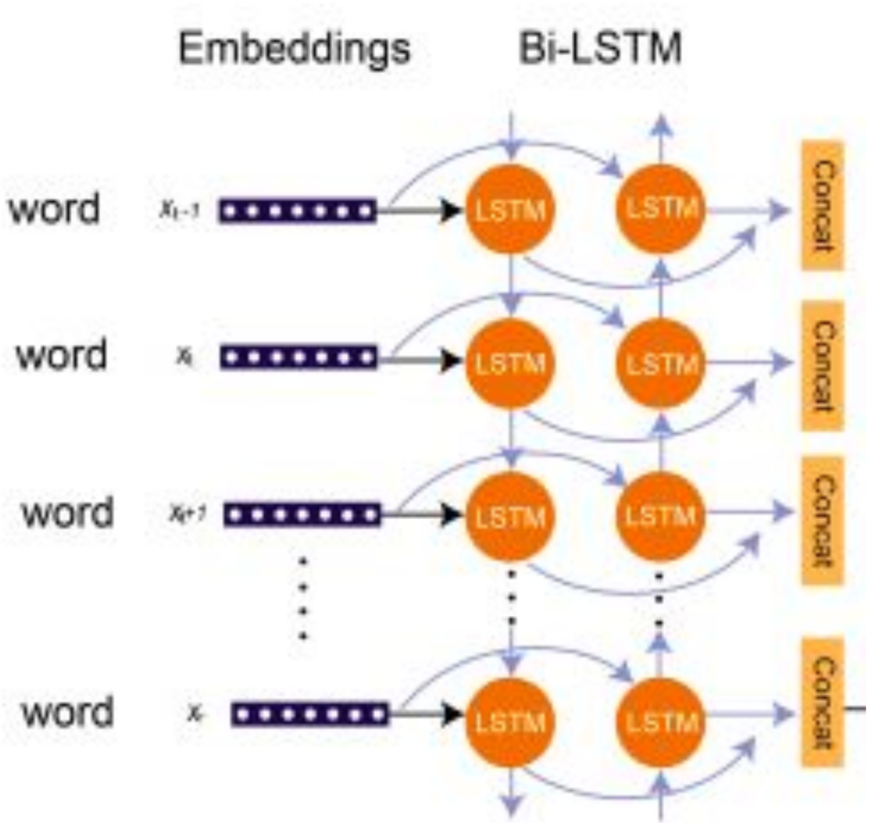
MAGNET network’s components :

- Feature Extraction
- Correlation between labels

Graph Attention Networks For Multi-label Classification

Feature Extraction

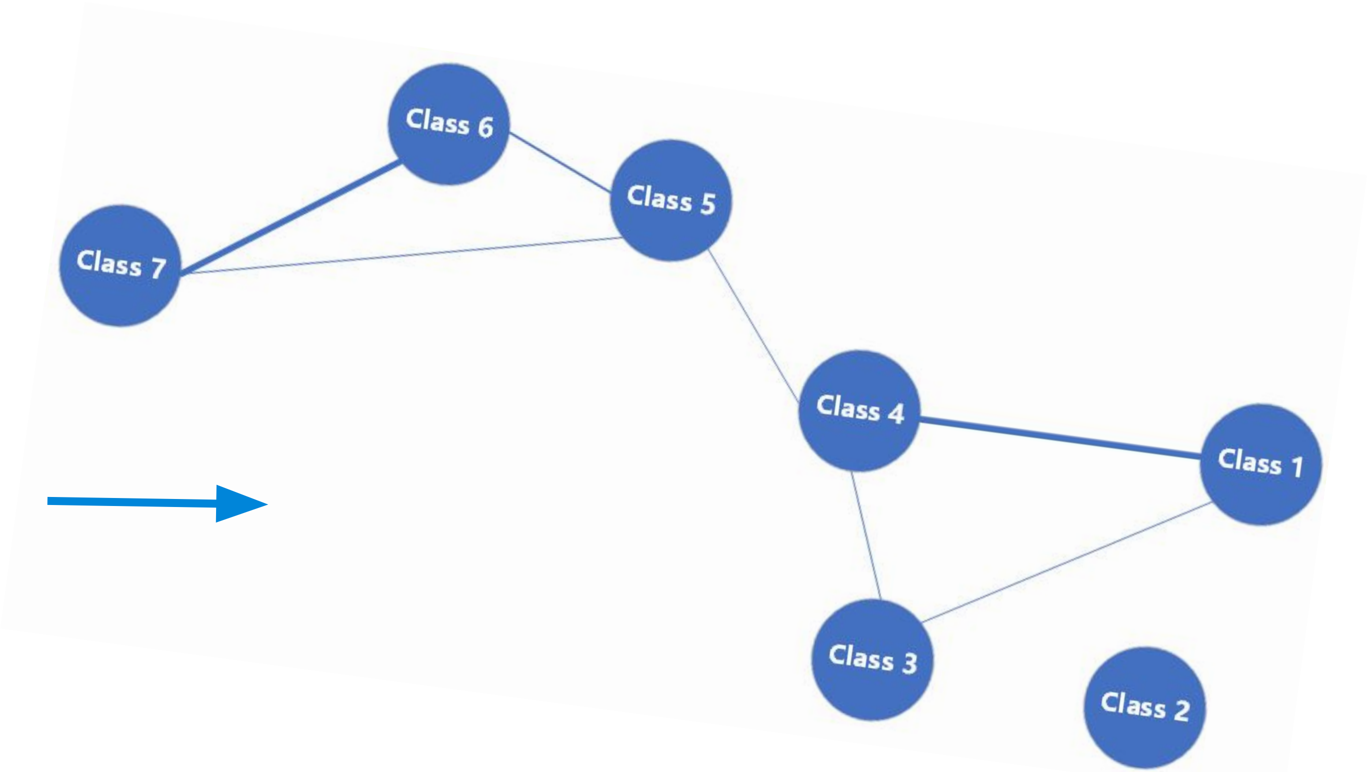
- Extracting Text features using LSTM network



Features

Reuters dataset

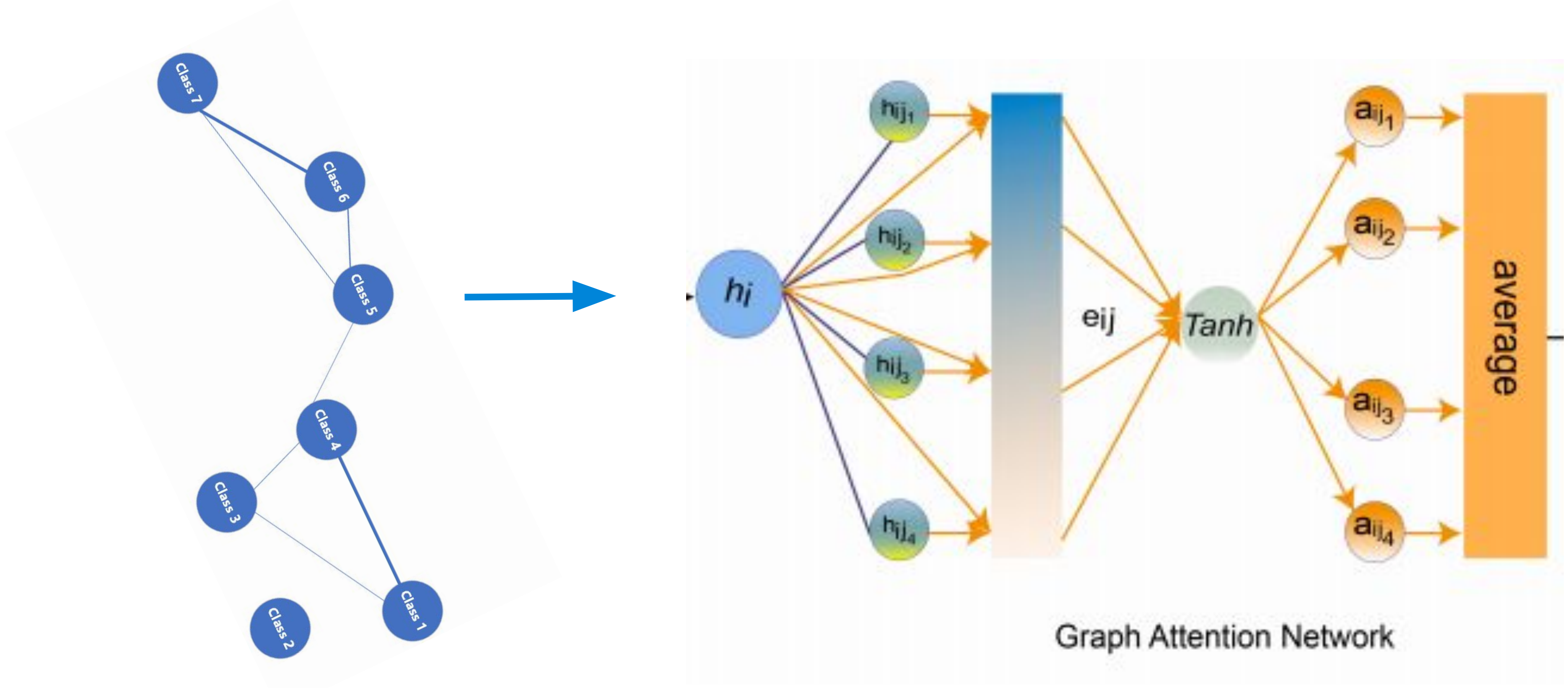
Graph is constructed by counting the pairwise co-occurrence of labels in dataset.



Graph Attention Networks For Multi-label Classification

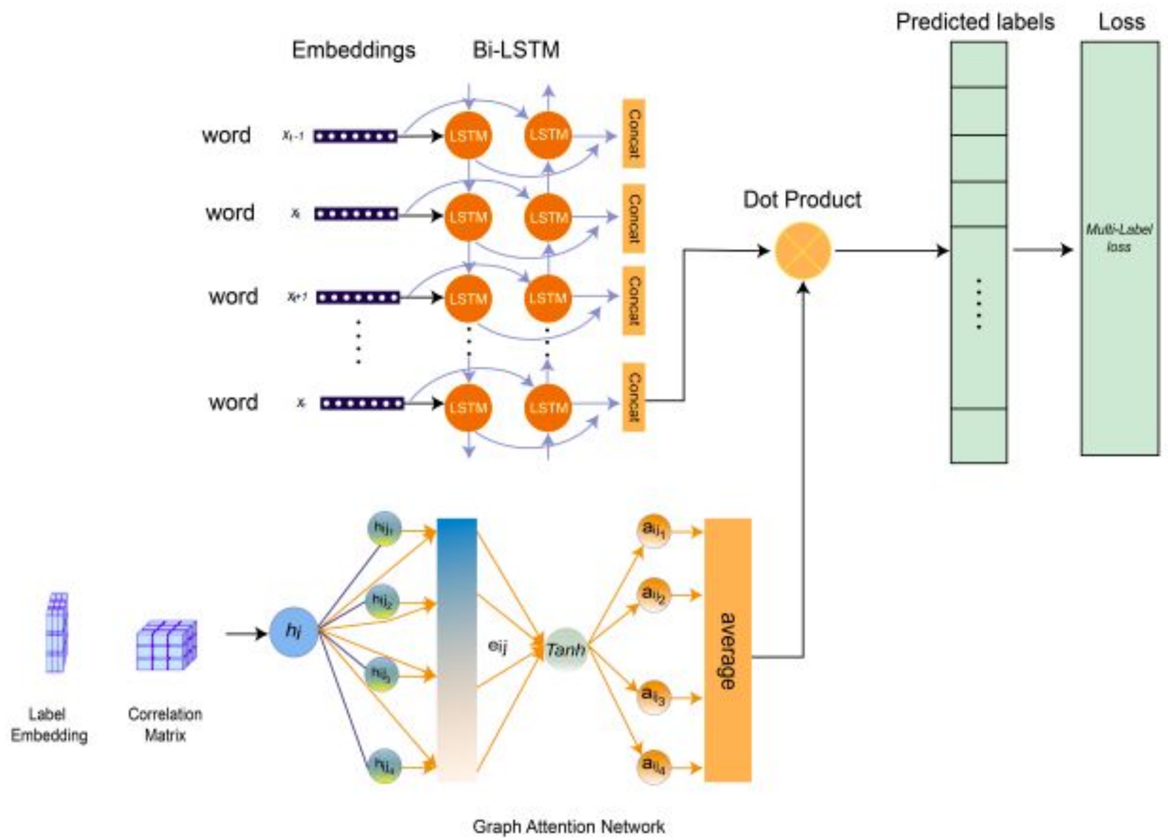
Correlation between labels

- Using Graph attention network to find correlation between labels



Graph Attention Networks For Multi-label Classification

Using both features to classify the documents



Graph Attention Networks For Multi-label Classification

Result

State-of-the-art accuracy compared to 27 different deep learning models including bert across five benchmark datasets

Table 3: Comparisons of Micro F1-score for various models on four benchmark datasets.

F1-accuracy				
Methods	**Reuters-21578**	**AAPD**	**Slashdot**	**Toxic**
BR	0.878	0.648	0.486	0.853
BR-support	0.872	0.682	0.516	0.874
CC	0.879	0.654	0.480	0.893
CNN	0.863	0.664	0.512	0.775
CNN-RNN	0.855	0.669	0.530	0.904
MAGNET	**0.899**	**0.696**	**0.568**	**0.930**

Table 4: Comparisons of Micro F1-score for various state-of-the-art models on Rcv1-v2 dataset.

Rcv1-v2	
Method	**Accuracy**
LR	0.692
SVM	0.691
HSVM	0.693
HLSTM	0.673
RCNN	0.686
XML-CNN	0.695
HAN	0.696
Bi-BloSAN	0.72
DCNN	0.732
SGM+GE	0.719
CAPSULE-B	0.739
CDN-SVM	0.738
HR-DGCNN	0.761
TEXTCNN	0.766
HE-AGCRCNN	0.778
BP-MLL_{RAD}	0.780
HTrans	0.805
BOW-CNN	0.827
HilAP	0.833
BERT	0.864
BERT + SGM	0.846
FMP + $LaMP_{pr}$	0.877
MAGNET	**0.885**

Take-Home messages

Graph neural networks :

- Graph neural networks captured better correlation compared to other deep learning models

- Learning the correlation matrix from scratch performed better than providing the pre-order correlation to Graph neural network

- The results of the proposed model are validated on five real-world MLTC datasets. The proposed model achieves similar or better performance compared to the previous state-of-the-art models.

Notes

Thanks !

aadityaura@gmail.com
twitter : @aadityaura