CHECK-IN #2

TweetGCN: A Graph Convolutional Network approach to topic classification on Twitter

Charles Duong (cduong5), Cecily Chung (cchung46), Benjamin Bradley (bpbradle)

Introduction

In this project, we propose an approach to topic classification on Twitter. We base our approach off of the TextGCN paper. TextGCN is an implementation of a GCN that casts GCNs to the field of NLP. The TextGCN paper addresses text classification and runs experiments on several datasets that classify long documents into topics.

For this project, we re-implement TextGCN in PyTorch and cast this to Twitter topic classification (making this a classification problem). The main difference that we will face is that Tweets are a lot shorter and the language is a lot more informal.

Related Work

Lu et al. published two related papers drawing on TextGCN for various Twitter-based applications. The first, SOSNet, introduced a cosine similarity score to construct a graph out of Tweets and an online dynamic query expansion to augment twitter datasets. The second, HateNet, introduced a stochastic probabilistic regularization technique to combat noisy network language and a substitution based dynamic query expansion technique to target data augmentation.

The following public implementations were used:

- Official implementation in Tensorflow: https://github.com/yao8839836/text_gcn

- Community implementation in PyTorch: https://github.com/codeKgu/Text-GCN

Data

We plan to use the CardiffNLP TweetTopic classification dataset which has literally thousands of tweets each assigned one of six labels which gives us the opportunity to teach our model how to sort the tweets based on their text into the correct category. This dataset is assembled by the Cardiff University NLP team for their 2022 NLP Hackathon and was later released on to the web for public use, the labels are relating to major topics of media such as Sports, Business, Pop Culture, etc.

Given that CardiffNLPs largest file has 4.37k Tweets each of which has its own label we hope to have enough data in this dataset to successfully pull off learning how to reliably sort the data by topic without having to preprocess.

We also have access to the larger but more complicated multi-topic dataset which is the same as the prior explanation except allows for multiple topics to be linked to one tweet rather than forcing it into only one topic. If our single topic dataset turns out to not be working or we simply want to expand the scope of the project we may turn to that: CardiffNLP Multi-topic TweetTopic

Methodology

The overall architecture of this paper would be:

1. Raw tweet data (straight from the dataset)
2. Preprocess data
3. Document embeddings
4. Graph construction
5. GCN layers
6. Tweet classification

Training formulation:

Let $X \in \mathbb{R}^{N \times F}$ represent our dataset, that is, X is the set of textual posts and let N be the number of posts in our input and let F be the number of features in our post.

We now construct a fully connected graph of online posts, $G = (V, E)$, where each post x_i is a node. Let each edge e be the cosine similarity between embeddings (the distance between the vectors).

To formulate our loss function, let $Y \in \{l_k | k = 1, 2, ..., K\}^N$, where l_k is the k^{th} tweet label (music, healthcare, tech, etc) in our labeled dataset and K is the number of classes.

We now formally define our problem as follows: given X and Y, we want to learn a function F paramaterized with W that maps X to Y: $F(X) \rightarrow Y$. We thus minimize the loss function as follows:

$$\arg\min_{E,W} \mathcal{L}(Y, F(X, W, E))$$

Challenges

The hardest part of implementing this paper we believe will be working around issues as they arise since GCNs are much less researched and talked about in a NLP context and we would have far less resources to lean on in case of emergency.

Metrics

Naively, we can test the accuracy of our TextGCN model on the dataset.

We can make our experiments more rigorous by implementing baseline models to test against. For example, we can implement a Linear regression, Naive Bayes, or SVM and compare our accuracy against those. We can also run McNemar's Tests or Student t to calculate a p score and see if our results are statistically significant.

The original TextGCN paper adapts a very similar approach of finding an accuracy and comparing it against baselines using the student t test.

Our null hypothesis (base goal) is that our TextGCN implementation makes no difference in the accuracy of classifying tweets. Our target goal is that our TextGCN implementation increases the accuracy of classifying tweets. Our alternative hypothesis (stretch goal) is that our TextGCN implementation makes a statistically significant difference in the accuracy of classifying tweets (calculated with the p score).

Ethics

1. The broader social implication of this paper is that as more of the social interaction happening in society moves online onto social media the importance of being able to accurately understand what is being talked about and precisely take down exactly those posts which fit the specific topics not allowed on platforms will become more and more important. Exploring another method for modeling differences in topics between posts on Twitter, a platform in the foreground of public conversation about hate speech and moderation, provides another tool in the toolbox of moderators and platform designers to build the social media platforms which we as a society need.

2. The major stakeholders relevant in this problem are the people making posts on social media, the people trying to read and enjoy their time on social media, and the companies which own these social media networks. Understanding our model as a text topic classifier, if it fails to perform accurately it will assign posts to the wrong topic which if posts are being recommended based on the topic assigned by our model would cause users to be receiving tweets they aren't interested in and/or cause posters to have their posts sent to the wrong audience and hurt their audience reach. Finally, social media companies have an interest in ensuring that posts which ought not be amplified or ought not be recommended to certain users be accurately not recommended so the accuracy of our model has a direct effect on the quality of social media recommendation algorithms if it were worked into that. A fourth stakeholder could potentially be

academics who seek to understand trends in conversation online and by improving our algorithm we ensure that researchers have a more realistic impression of conversation online.

Division of Labor

All three members have contributed equally on planning, coding, and reporting in our project.