Zero Shot Learning for Text Classification

Er.Romika Manhas[1], Dr. Simmi Dutta[2], Dr. Jyoti Kumar Mahajan[3]
[1]*P.G. student, Dept. of Computer Engineering, Govt. College of Engg. & Technology, Jammu, India*
[2]*Head of Dept., Dept of Computer Engineering, Govt. College of Engg. & Technology, Jammu, India*
[3]*Professor, Department of Computer Engineering, Govt. College of Engg. & Technology, Jammu, India*

Abstract-Zero shot learning aims at solving the problem of object classification where we have no training examples as insufficient and uneven datasets for emerging classes is becoming a challenging issue monotonously. Zero shot text classification is an approach where we use zero shot learning to make models which are capable of categorizing the text documents which it has not seen during the testing phase. In this paper we will be giving a brief idea about zero shot learning and text classification with the different state-of- art methods which has been used for zero shot text classification problem which includes neural network with embeddings, semantic embeddings, using knowledge graphs etc. In addition we have applied distil BERT -a pretrained model, for seen classes and the results for the same has been discussed.

Keywords: zero shot learning, zero shot text classification, semantic embeddings, BERT, knowledge graphs.

I.INTRODUCTION

Automatic classification has been one of the most fundamental problems in object classification under machine learning. Machines are only capable of recognizing hundreds and thousands of classes and are unable to classify objects which are unseen to it ,however in contrast, according to a study human are capable of distinguish between 3000 basic object categories with addition to more subordinate ones. To free recognition tasks from collecting of large labelled image datasets, zero-shot learning (ZSL) is gaining an increased attention in recent years, which aims to recognize instances from the new unseen categories. With the label sets between seen and unseen categories being disjoint, the key in the general methodology of ZSL is to establish the interclass connections via intermediate semantic representations, either manually defined by human experts annotated attributes or automatically extracted from auxiliary text sources. However, many approaches, proven to be effective in traditional classification tasks, cannot catch up with a dynamic and open environment where new classes can emerge after the learning stage. For example, the number of topics on social media is growing rapidly, and the classification models are required to recognize the text of the new topics using only general information (e.g., descriptions of the topics) where labelled training instances are unfeasible to obtain for each new topic. This scenario holds in many real-world domains such as medical diagnosis and object detection. To eliminate these challenges researchers are now focusing on different methods for implementing zero shot learning. In this paper, 1) we have discussed about different approaches applied for implementing ZSL for the ‘text classification’ problem 2) we have given our approach for seen classes 3) we have discussed the direction of future developments

II.LITERATURE SURVEY

One of the earliest work in this field is the data less classification models [1] [2] in 2014. Here the model of classification does not need any annotated training data. The approach is based on the usage of a source of world knowledge to analyze both labels and documents from a semantic point of view. One of the approaches that has been used by Pushpankar Kumar Pushp & Muktabh Mayank Srivastava [4] in 2017 is word embedding. This approach focused on making the model learn about the relationship between a sentence and embedding of sentence’s tags. Learning such relationship makes the model generalize to unseen sentences, tags, and even new datasets provided they can be place into same embedding space. The model learns to predict whether or not a given sentence is said to a tag or not.
Integrating semantic knowledge from knowledge graph is one of the remarkable approach used for ZSL. This was used in concept ontology, that included class hierarchy and knowledge graphs, which represents relationships among classes and features, by Wang et al.,

2018 and Fergus et al., 2010[5][6]. This concept of integrating knowledge graph for ZSL has been used for text classification by Jingqing et al [7] in 2019 by proposing a two phase model which uses ConceptNet knowledge graph and called it feature augmentation in addition to data augmentation. In another approach used by Wenpeng Yin et al [8], standardization of dataset was done in order to achieve the goal and provided datasets for detecting emotions as well. Similar approaches where the dataset was considered instead of model was by meta tuning datasets and was introduced by Ruiqi Zhong et al in 2021[9]. In another approach by Yu Zhang1, Zhihong Shen in 2022[10], metadata induced learning was introduced in order to use for ZSL and called it MlCOl. These approaches has remarkable results but work with a specific datasets. In June 2022, Qi Chen , Wei Wang et al [11] introduced an approach for knowledge graph embedding using the pretrained model called BERT by projecting BERT space embedding with the pre-existing knowledge graph embedding space.

III.METHODOLOGY OF PROPOSED WORK

Problem definition:

In ZSL we want our model to classify text not only from seen but also unseen classes. More accurately, we can define it as: Given labeled instances belonging to a set of seen classes S, ZEROSHOT-TC aims at learning a classifier $f(\cdot) : X \rightarrow Y$, where $Y = S \cup U$; U is a set of unseen classes and belongs to the same aspect as S.

Dataset used:

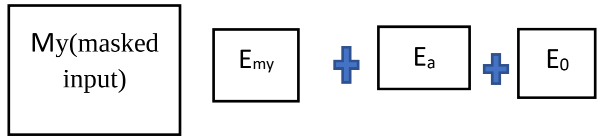
The dataset used here is 20 newsgroups text dataset which has been imported from sklearn. It comprised of 18000 newsgroup posts on 20 topics.

Model used:

In this approach for seen classes distl BERT [14] has been used which is the lighter version of BERT[15].BERT stands for Bidirectional Encoder Representations from Transformers. It is a pre-trained model that can be finetuned with just one additional output layer to create state-of-the-art models for a wide range of tasks without any task specific architecture modifications.

Figure1: input embeddings are given as the sum of below embeddings i.e

1. token embeddings
2. segmentation embeddings
3. position embeddings



In distil BERT compression of BERT has been done and it has been reduced by 40% and being 60% faster than BERT. It has the same general architecture as BERT except the token-type embeddings and the pooler are removed while the number of layers is reduced by a factor of 2.

IV.RESULT

The result has been represented for seen classes in the form of graph. It shows precision, recall and f1 score. It can be seen that the max achieved precision for a particular class is .99 that is 99%. The max recall is 97%. The average result has been given in the tabular form below with average accuracy, precision and recall.

Figure2: Graph of precision, recall and f1score for seen classes.

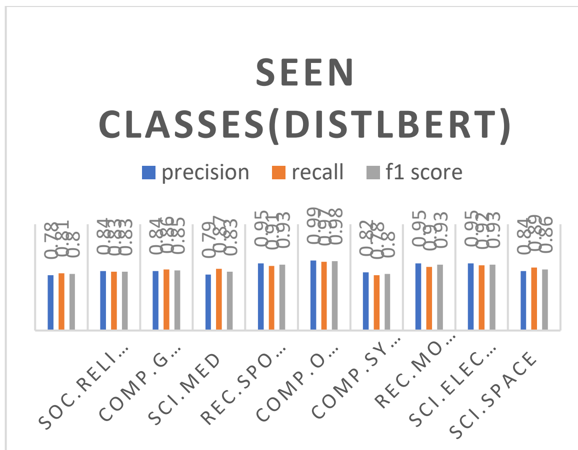


Table 1: Average accuracy, recall and f1 score of the model.

	precision	recall	F1 score
Accuracy			.87
Macro avg	.87	.87	.87
Weighted avg	.88	.87	.87

V.CONCLUSION AND FUTURE SCOPE

The study shows that BERT is indeed an effective choice to be used as a classifier for especially for seen classes. The overall accuracy is upto 88% which we have achieved by using its variant distl BERT. Some its variant such as BERT KG, s-BERT can be used for future works for unseen classes. In the future, we plan to extend our work by doing the classification with a larger amount of data, and for unseen classes.

REFERENCE

[1] Y. Song, D. Roth, On dataless hierarchical text classification, in: Proceedings of the 28th AAAI conference on Artificial Intelligence, 2014

[2] M. W. Chang, L. Ratinov, D. Roth, V. Srikumar, Importance of semantic representation: Dataless classification, 27th AAAI conference on Artificial Intelligence (2008).

[3] Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach Wenpeng Yin, Jamaal Hay, Dan Roth Cognitive Computation Group Department of Computer and Information Science, University of Pennsylvania.

[4] Train once, test anywhere: zero-shot learning for text classification pushpankar Kumar Pushp & Muktabh Mayank Srivastava ParallelDots, Inc.

[5] Xiaolong Wang, Yufei Ye, and Abhinav Gupta. 2018. Zero-shot recognition via semantic embeddings and knowledge graphs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 6857–6866

[6] Rob Fergus, Hector Bernal, Yair Weiss, and Antonio Torralba. 2010. Semantic label sharing for learning with many categories. In Computer Vision – ECCV 2010, pages 762–775, Berlin, Heidelberg. Springer Berlin Heidelberg.

[7] Integrating Semantic Knowledge to Tackle Zero-shot Text Classification Jingqing, Piyawat Lertvittayakumjorn ,Yike Guo Data Science Institute Imperial College London London, UK

[8] Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach Wenpeng Yin, Jamaal Hay, Dan Roth Cognitive Computation Group Department of Computer and Information Science, University of Pennsylvania

[9] Adapting Language Models for Zero-shot Learning by Meta-tuning on Dataset and Prompt Collections Ruiqi Zhong Kristy Lee∗ Zheng Zhang∗ Dan Klein Computer Science Division, University of California, Berkele

[10] Metadata-Induced Contrastive Learning for Zero-Shot Multi-Label Text Classification Yu Zhang1 , Zhihong Shen2 , Chieh-Han Wu2 , Boya Xie2 , Junheng Hao3 , Ye-Yi Wang2 , Kuansan Wang2 , Jiawei Han1

[11] Zero-Shot Text Classification via Knowledge Graph Embedding for Social Media Data Qi Chen , Wei Wang , Kaizhu Huang , Senior Member, IEEE, and Frans Coenen.

[12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In NAACL-HLT, 2018.

[13] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019

[14] DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter by Victor SANH, Lysandre DEBUT, Julien CHAUMOND, Thomas WOLF Hugging Face

[15] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova Google AI Language