CIS519 Project Description

November 2020

1 Introduction

In this final project for CIS 519, you will work in groups of 2-4 students to design and conduct a large scale machine learning experiment using the concepts you have learned in class and those that you will acquire from the relevant literature, as you design and work on the project.

We are providing three project options for you to choose from, spanning three input modalities: Natural Language Text, Images, and Audio. In each case, we also provide datasets that will constitute (some of) the training data you will use, and the test data on which you will evaluate the performance of your learned models. However, as we describe below, the machine learning components of the proposed projects will have some commonalities.

The overarching theme across all projects is *zero-shot learning* in the context of multi-class classification. In each case, the goal is to classify objects into one of l different classes. However, you will be given training data only for some of these labels, and will have to figure out way to classify examples that belong to new labels, that you have not observed in the training data.

There is a lot of literature on zero-shot learning that you can consult, and we hope that you can bring up some new ideas. In the domain of natural language, research on zero-shot learning has started with this work [2] (which was called *Dataless classification*, before zero-shot classification became the agreed upon term). This was followed up by works on zero-shot classification in multilingual classification [13] and many other tasks including semantic typing [15], Relation extraction [9], and others. In most cases methods make use of external information sources such as Wikipedia, or transfer from other tasks [14].

Zero-shot classification has been use a lot in Computer Vision too, starting with [7], followed by a very rich body of work, including [12, 1] that focused on learning representations.

Below we provide information on the datasets and the general idea of each project. Your job is to design and implement the project and, of course, try to contribute new ideas as much as you can.

In all cases, we will provide the datasets, dataset splits, and also set up leader boards for some of the tasks, so that you will be able to gauge your progress.

2 Timeline

The timeline for the project is:

11/11: choose a project and form teams

11/23: submit initial proposals describing what your team plans to do

12/2: progress report

12/15-20: (TBD) final paper + short video due

3 Option 1: Computer Vision

Within the computer vision domain, object classification is a task that has been one of the greatest demonstrators of the power of deep learning. MNIST, released in 1998, is regarded as the "hello world" of the computer vision domain, and now in 2020 the task of classifying this dataset is largely regarded as 'solved'.

As it became clear that our classifiers are powerful and can learn to model large enough annotated datasets, interest is beginning to shift to deal with the more realistic cases where we can do not have annotated data for all tasks of interest. Zero-shot learning addresses one important aspect of this challenge. In this project you will study this important problem in the context of image classification, and study ways to classify objects into unseen classes, i.e. objects types that did not appear in training images. Most Zero-Shot approaches in computer vision rely on representation learning, assuming that good enough representations can be induced while learning to classify other types of objects and, sometimes, also on some form of text data. One rather general pipeline used for this process could be the following two step process:

- You run a CNN on the image, learning/predicting 'attributes' of your images
- You classify the image based on which attributes the CNN says are present, whether through manual definition of these attributes, or learning them through utilization of text data

This way you can classify classes that aren't in your training set. For example, imagine your model has seen pictures of zebras, and is therefore able to identify the attribute 'striped', it has also seen hamsters and therefore able to identify the attribute 'fuzzy'. Now, at test time, if you see a bee and bees were not present in your training images, the model will still be able to predict 'striped' and 'fuzzy' in step 1 of the above pipeline, and then in step 2 you can use the fact that bees are fuzzy and striped to guess that the image is of a bee. Note that to use this approach, while you don't need to have seen images of these zero-shot classes, you will need information about these classes and their attributes in order to classify them.

Other approaches exist that rely on first *pre-training* representations on existing annotated datasets, or on the textual definitions of the target labels.

3.1 Dataset

The dataset that we will be using for this task is the Caltech UCSD bird dataset, one of the most widely used and competitive fine-grained classification benchmarks. You can further investigate the data at this link. You will be provided with the following:

- A dataset consisting of 175 bird species

- A set of txt files identifying attributes that are present in each of the images that you are provided

While you are given the attributes that are present in each image, you are not given descriptions of the classes (e.g. you will be told that image 2050 has a yellow beak, but you do not know which birds in your dataset have yellow beaks). We encourage the use of outside information, specifically you will want to look into obtaining descriptions of these different bird species for step two in the above pipeline (you could scrape wikipedia for example). You are NOT allowed to use the images of the 25 birds that are not present in the data we give you, however you will be given the names of these zero-shot birds, as you will need this for obtaining descriptions of them for classification.

3.2 Suggested Reading

Some of the relevant references [7, 12, 1] are listed below.

- Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer
- An embarrassingly simple approach to zero-shot learning
- Predicting Deep Zero-Shot Convolutional Neural Networks using Textual Descriptions (slightly more advanced)

4 Option 2: Natural Language Processing

As pointed out above, Zero-Shot was first studied in the context of natural language text classification. However, in this project, you will work on another challenging problem, that of Named Entity Recognition (NER). The task in this case is to identify what phrases of text constitute a named entity, and classify it into one of 18 different types. However, the challenge will be that you will be given annotated data for only some of the labels, and you will need to identify also named entities of different types, that were not annotated in the training data. That is, you will evaluate your models on their ability to identify named entities and classify them into all 18 labels.

We will set up several training datasets; in each one, a different subset of the labels will be annotated. For example, we will have training sets with 12 annotated labels, 6 annotated labels, and no annotated labels.

In your work you can use existing approaches to NER such as neural approaches [8, 5], or non-neural approaches as in [11, 4, 6].

You may also consider pre-training of embeddings, such as BERT [3] or RoBERTa [10], and the use of external resources as in [15].

4.1 Dataset

The dataset we will use is OntoNotes 5.0. OntoNotes is a large dataset with annotations for several tasks in three languages. We will only use the annotation of NER in English. You do not need to download this dataset, or write a reader by yourself. We will provide the preprocessed data with the appropriate data splits. For example:

- For the model trained with 6 labels, the 6 labels you will use is: PERSON, ORGANIZATION, GPE, EVENT, NORP, TIME.

- For the model trained with 12 labels, the 12 labels you will use is: PERSON, ORGANIZATION, GPE, EVENT, NORP, TIME, FACILITY, LAW, PERCENT, QUANTITY, WORK_OF_ART, PRODUCT.

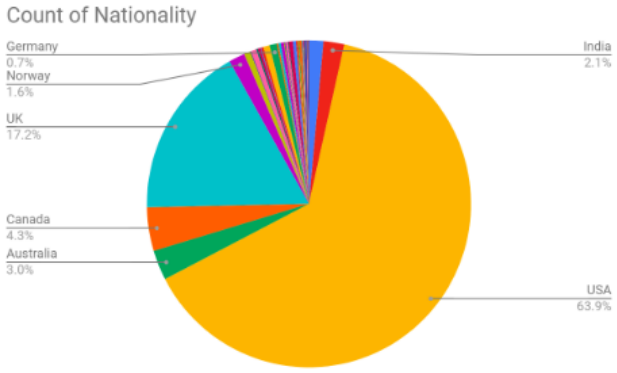
5 Option 3: Audio Classification

In this project, you will be asked to classify audio speech samples by gender and nationality. However, instead of simply identifying the gender and nationality of a voice in an audio clip, you will take a zero-shot learning approach. You will train a model to identify the nationality of a speaker, using only data from a subset of the entire set of nationalities. Then you will predict nationality of a speaker from the entire set of nationalities, including those not seen during training. Your goal will be to improve performance on these unseen nationalities. Use any zero-shot learning tricks mentioned in readings or found online.

The tasks:

1. Train on all English-speaking nationalities (e.g. USA, UK, Ireland, Australia, Canada, etc.) and evaluate performance on non English-speaking nationalities.
2. Train on just the USA subset of samples, then evaluate performance on the rest of the English-speaking nationalities.
3. Train a classifier for nationalities, only using the male speakers. Evaluate performance of the classifier on the female speakers.

See the distribution of nationalities below:



Since with audio files, the signals are inherently noisy, you will need to consider preprocessing and featurization steps to transform the raw audio samples into data you can use in your model. Here are some suggestions and pointers to start with:

- Processing audio data for machine learning: Youtube link
- Removing noise from a signal: Blog post
- Spectrograms: Blog post

5.1 Dataset

The dataset that we will be using for this task is the VoxCeleb1 dataset. VoxCeleb1 is already split into development and test sets. Each datapoint has an id pointer to which audio clip it corresponds to, and the nationality and gender of the primary speaker in that audio clip. To obtain the data, find the section 'Audio files.' You will need to request and download the audio data. The metadata for all speakers in the dataset can be found in the section 'Metadata.'

References

[1] L. J. Ba, K. Swersky, S. Fidler, and R. Salakhutdinov. Predicting deep zero-shot convolutional neural networks using textual descriptions. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 4247–4255. IEEE Computer Society, 2015.

[2] M.-W. Chang, L. Ratinov, D. Roth, and V. Srikumar. Importance of Semantic Representation: Dataless Classification. In *Proc. of the Conference on Artificial Intelligence (AAAI)*, 7 2008.

[3] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *North American Association for Computational Linguistics (NAACL)*, pages 4171–4186, Minneapolis, Minnesota, June 2019.

[4] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*, Ann Arbor, Michigan, June 2005. Association for Computational Linguistics.

[5] Z. Huang, W. Xu, and K. Yu. Bidirectional LSTM-CRF models for sequence tagging. *CoRR*, abs/1508.01991, 2015.

[6] D. Khashabi, M. Sammons, B. Zhou, T. Redman, C. Christodoulopoulos, V. Srikumar, N. Rizzolo, L. Ratinov, G. Luo, Q. Do, C.-T. Tsai, S. Roy, S. Mayhew, Z. Feng, J. Wieting, X. Yu, Y. Song, S. Gupta, S. Upadhyay, N. Arivazhagan, Q. Ning, S. Ling, and D. Roth. CogCompNLP: Your Swiss Army Knife for NLP. In *Proc. of the International Conference on Language Resources and Evaluation (LREC)*, 2018.

[7] C. H. Lampert, H. Nickisch, and S. Harmeling. Learning to detect unseen object classes by betweenclass attribute transfer. In *In CVPR*, 2009.

[8] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer. Neural architectures for named entity recognition. *CoRR*, abs/1603.01360, 2016.

[9] O. Levy, M. Seo, E. Choi, and L. Zettlemoyer. Zero-shot relation extraction via reading comprehension. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 333–342, Vancouver, Canada, Aug. 2017. Association for Computational Linguistics.

[10] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. RoBERTa: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[11] L. Ratinov and D. Roth. Design Challenges and Misconceptions in Named Entity Recognition. In *Proc. of the Conference on Computational Natural Language Learning (CoNLL)*, 6 2009.

[12] B. Romera-Paredes and P. Torr. An embarrassingly simple approach to zero-shot learning. volume 37 of *Proceedings of Machine Learning Research*, pages 2152–2161, Lille, France, 07–09 Jul 2015. PMLR.

[13] Y. Song, S. Upadhyay, H. Peng, S. Mayhew, and D. Roth. Toward any-language zero-shot topic classification of textual documents. *Artificial Intelligence*, 274:133–150, 2019.

[14] W. Yin, J. Hay, and D. Roth. Benchmarking Zero-shot Text Classification: Datasets, Evaluation, and Entailment Approach. In *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019.

[15] B. Zhou, D. Khashabi, C.-T. Tsai, and D. Roth. Zero-Shot Open Entity Typing as Type-Compatible Grounding. In *Proc. of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2018.