SC4001 CE/CZ4042: Neural Networks and Deep Learning

Group Project

Last Update: 14 Mar 2025

Start Date: 14 Mar 2025 Deadline: 11 April 2025 (11:59 PM)

Students are to propose and execute a final project on an application or a research issue that is related to neural networks and deep learning. The project can be carried out in a group consisting of no more than three members. Students are to come up with a potential technique for the application or to mitigate the issue, to develop associated codes, and to compare with existing methods. Students may choose, focus, and expand on the project ideas **A – F** given below.

By the deadline, students are to submit a project report in a .pdf file of **ten A4 pages** (Arial 10 font) and associated **code** in a .zip file to NTULearn.

The project report should have the names of the team members on the front page and contain an introduction to the project idea, a review of existing techniques, a description of the methods used, experiments and results, and a discussion. The 10-page limit is exclusive of references, content page, and cover page. The code needs to be commented properly. Make sure the code can be tested easily.

The assessment is based on the project execution (30%), experiments and results (30%), report presentation (15%), and novelty (15%), and peer review (10%. Conducted via Eureka). We apply the same late submission penalty as in Assignment 1, i.e., 5% for each day up to three days.

A. Speech Emotion Recognition

Speech emotion recognition (SER) is the prediction of speaker's emotions from speech signals. SER involves extraction of audio features from speech and classification of speaker utterances to emotional classes.

Interesting projects would be

1. To develop deep learning techniques for SER invariant to speaker characteristics such as gender, age emotion.
2. To develop unsupervised learning techniques for SER
3. To detect emotions dynamically in speech. That is, to predict emotions within subintervals of the speech utterance.

References:

1. S. Zhang, S. Zhang, T. Huang and W. Gao, "Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid Matching," in *IEEE Transactions on Multimedia*, vol. 20, no. 6, pp. 1576-1590, June 2018
2. K. Han, D. Yu, and I. Tashev, "Speech emotion recognition using deep neural network and extreme learning machine," in Proceedings of Interspeech, 2014.

Datasets:

1. EMO-DB: https://www.kaggle.com/datasets/piyushagni5/berlin-database-of-emotional-speech-emodb/
2. RAVDESS: https://www.kaggle.com/uwrfkaggler/ravdess-emotional-speech-audio

For audio feature extraction:

1. Opensmile: https://www.audeering.com/opensmile/

B. Text Emotion Recognition

Text emotion recognition (TER) involves predicting emotions expressed in text and documents. Existing algorithms find emotion by learning the relationships of words using recurrent neural networks (RNN) or convolutional neural networks (CNN). RNN and CNN capture local information (i.e., emotion of words) and ignore the global information (i.e., emotion of sentence).

Interesting projects would be

1. To develop deep learning techniques for capture both local and global information. The local information refers to emotions expressed by words and global information refers to emotions expressed by the meanings of sentences.
2. To develop techniques that are invariant to speaker's writing styles and characteristics

References:

1. Kim, Y. Convolutional neural networks for sentence classification, Conference on Empirical Methods in Natural Language Processing, pp. 1746–1751, 2014.
2. Lai, S., Xu, L., Liu, K., & Zhao, J. Recurrent convolutional neural networks for text classification. Proceedings of the National Conference on Artificial Intelligence, vol. 3, pp. 2267–2273, 2015.
3. Zhou, P., Qi, Z., Zheng, S., Xu, J., Bao, H., & Xu, B. Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling. 26th International Conference on Computational Linguistics, vol. 2, no.1, 3485–3495, 2016.

Datasets:

1. CROWDFLOWER: https://data.world/crowdflower/sentiment-analysis-in-text
2. WASSA2017: https://github.com/vinayakumarr/WASSA-2017/tree/master/wassa

C. Sentiment Analysis

Text sentiment analysis (TSA) refers to identification of sentiments, usually positive or negative, expressed in text or document. One may want to develop deep learning techniques for TSA

1. To deal with domain adaptation, that is, how can one adapt a network train on one domain to work in another domain
2. To compare the performance of different Transformers architectures
3. To deal with small datasets, that is, with insufficient number of training samples

References:

1. T. Gui *et al.*, "Long Short-Term Memory with Dynamic Skip Connections," *Proc. AAAI Conf. Artif. Intell.*, 2019, doi: 10.1609/aaai.v33i01.33016481.
2. A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *ACL-HLT 2011 - Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011.
3. X. Zhang, J. Zhao, and Y. Lecun, "Character-level convolutional networks for text classification," in *Advances in Neural Information Processing Systems*, 2015.

Datasets:

1. Stanford Sentiment Treebank: https://www.kaggle.com/atulanandjha/stanford-sentiment-treebank-v2-sst2
2. IMDB movie review dataset: https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews
3. YELP review dataset: http://xzh.me/docs/charconvnet.pdf

D. Gender Classification

Automatic gender classification has been used in many applications. The goal of this project is to classify the gender of faces in an image. One can design a convolutional neural network or Transformer to achieve this goal. Some tasks to consider:

1. Modify some previously published architectures e.g., increase the network depth, reducing their parameters, etc. Explore more advanced techniques such as deformable convolution, dilated convolution (dilation>1) or visual prompt tuning for Transformers.
2. Consider age and gender recognition simultaneously to take advantage of the gender-specific age characteristics and age-specific gender characteristics inherent to images
3. Consider pre-training using the CelebA dataset http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html

References

1. G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks." in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) workshops*, 2015
2. Z. Liu and P. Luo and X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *International Conference on Computer Vision (ICCV),* 2015

Datasets:

1. Adience Dataset: https://talhassner.github.io/home/projects/Adience/Adience-data.html#agegender

E. Clothing Classification

Fashion-MNIST is a dataset of Zalando's article images—consisting of a training set of 60,000 examples and a test set of 10,000 examples. Each example is a 28x28 grayscale image, associated with a label from 10 classes. One can design a convolutional neural network or Transformer to address the classification problem. Some tasks to consider:

1. Modify some previously published architectures e.g., increase the network depth, reducing their parameters, etc. Explore more advanced techniques such as deformable convolution, dilated convolution (dilation>1) or visual prompt tuning for Transformers.
2. Use more advanced transformation techniques such as MixUp (see the original paper and its PyTorch implementation here)
3. Comparing the performance of different network architectures

References

3. Deep Learning CNN for Fashion-MNIST Clothing Classification

Datasets:

1. The dataset is available in TorchVision https://pytorch.org/vision/stable/generated/torchvision.datasets.FashionMNIST.html

F. Flowers Recognition

The Oxford Flowers 102 dataset is a collection of 102 flower categories commonly occurring in the United Kingdom. Each class consists of between 40 and 258 images. The images have large scale, pose and light variations. In addition, there are categories that have large variations within the category and several very similar categories.

The dataset is divided into a training set, a validation set and a test set. The training set and validation set each consist of 10 images per class (a total of 1020 images each). The test set consists of the remaining 6149 images (minimum 20 per class). Some tasks to consider:

1. Modify some previously published architectures e.g., increase the network depth, reducing their parameters, etc. Explore more advanced techniques such as deformable convolution, dilated convolution (dilation>1) or visual prompt tuning for Transformers.
2. Analyze the results of using fewer training images, i.e., few-shot learning
3. Use more advanced transformation techniques such as MixUp (see the original paper and its PyTorch implementation here)
4. Try more advanced loss function such as triplet loss

References:

1. Nilsback, M-E. and Zisserman, A., “Automated flower classification over a large number of

classes," *in British Machine Vision Conference (BMVC)*, 2008

Datasets:

1. The dataset is available in TorchVision https://pytorch.org/vision/main/generated/torchvision.datasets.Flowers102.html
2. The Oxford Flowers 102 Dataset https://www.robots.ox.ac.uk/~vgg/data/flowers/102/

Computational Resource

You can use the computational resources assigned by the course. Alternatively, you can use Google Colab for computation. Note that the free version of Colab has a session duration limit, after which the environment needs to be reset. This can be disruptive for long-running experiments or processes.