CE/CZ4042: Neural Networks and Deep Learning

Guidelines for the Project

Deadline: 6 November 2020

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. 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. The project can be carried out in a group comprised of no more than three members. Project groups are to be formed by students themselves.

By the deadline, students are to submit a project report in a .pdf file, associated codes in a .zip file, and a video highlighting the project in .mpg format to NTU learn. The project report should have the names of the team members on the front page and would 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 codes need to be commented properly and easily tested. The video presentation is typically less than 10 minutes and highlights the methods and major achievements of the project.

The assessment is on based on the project execution (30%), experiments and results (30%), presentation (20%), and novelty (20%).

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 such as gender, age, accent, etc.
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: http://emodb.bilderbar.info/start.html
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 emotions 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 one can adapt a network train in one domain to work in another domain
2. To avoid using recurrent networks in order to speed up computations
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 including image analysis on social platforms. The goal of this project is to classify the gender of faces in an image. One can design a convolutional neural network to achieve this goal. Some tasks to consider:

1. Modify some previously published architectures e.g., increase the network depth, reducing their parameters, etc.
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. Material Recognition

The goal of this project is to train a convolutional neural network to classify color photographs of surfaces into one of ten common material categories: fabric, foliage, glass, leather, metal, paper, plastic, stone, water, and wood. Some tasks to consider:

1. Modify some previously published architectures e.g., increase the network depth, reducing their parameters, etc.
2. Try data augmentation to increase the number of training images
3. Try a larger dataset, Materials in Context Database (MINC)

References

1. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz, "Exploring features in a Bayesian framework for material recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 2010
2. S. Bell, P. Upchurch, N. Snavely, K. Bala, "Material recognition in the wild with the materials in context database," *in IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015

Datasets:

1. Flickr Material Database (FMD): https://people.csail.mit.edu/lavanya/fmd.html

F. Car Make Recognition

Automated car model analysis, particularly the fine-grained car categorization and verification, can be used for innumerable purposes in intelligent transportation system including regulation, description and indexing. For instance, fine-grained car categorization can be exploited to inexpensively automate and expedite paying tolls from the lanes, based on different rates for different types of vehicles. The goal of this project is to classify an image into one of the 163 car makes. Some tasks to consider:

1. Modify some previously published architectures e.g., increase the network depth, reducing their parameters, etc.
2. Consider a multi-task learning framework that classifies not only car makes, but also car models and the corresponding attributes
3. Try more advanced loss function such as triplet loss

References:

1. L. Yang, P. Luo, C. C. Loy, X. Tang, "A large-scale car dataset for fine-grained categorization and verification," *in IEEE Computer Vision and Pattern Recognition (CVPR)*, 2015
2. H. Liu, Y. Tian, Y. Wang, L. Pang, T. Huang, "Deep relative distance learning: tell the difference between similar vehicles," *in Computer Vision and Pattern Recognition (CVPR)*, 2016

Datasets:

1. The Comprehensive Cars (CompCars) Dataset
   http://mmlab.ie.cuhk.edu.hk/datasets/comp_cars/index.html