FYP: Deep Speaker Embedding Across Languages

This project aims to incorporate language-mismatch loss and contrastive loss into the a speaker embedding network for language-independent speaker verification. The student will use PyTorch and Python to implement the network and use Mandarin and English speech data for training and evaluation.

1. Python:
https://www.learnpython.org

2. Development Environments:

- Install Linux or use cloud services (e.g. https://colab.research.google.com, AMS, Ali Cloud, etc.)
- Install Anaconda with Python 3.9 and create an Anaconda environment. Read the documentation of Anaconda to know how to manage an Anaconda environment.
- Install the following package on the Anaconda environment that you have created.
    - pytorch
    - scipy
    - scikit-learn
    - matplotlib
    - librosa
- Install IDE: VS Code, Spyder, or Pycharm (I found VS Code the best)

3. DNN Platforms

- PyTorch
- https://pytorch.org/tutorials/beginner/deep_learning_60min_blitz.html
- https://pytorch.org/tutorials/beginner/pytorch_with_examples.html
- https://pytorch.org/tutorials/recipes/recipes_index.html

4. Speaker Verification: Book, Data and Software

a) Download Voxceleb1 and Voxceleb2 (http://www.robots.ox.ac.uk/~vgg/data/voxceleb/). If the datasets are not available there, try http://bioinfo.eie.polyu.edu.hk/download/voxceleb1.tgz and http://bioinfo.eie.polyu.edu.hk/download/voxceleb2.tgz
Decompress the file by "tar zxvf voxceleb1.tgz"
Download CN-Celeb1 and CN-Celeb2 from http://cnceleb.org and read the corresponding papers

b) Read this book: M.W. Mak and J.T. Chien, *Machine Learning for Speaker Recognition*, Cambridge University Press, 2020. [You may find this book in PolyU eLibrary]

c) Read the paper: H.Q. Mao, F. Hong, and M.W. Mak, "Cluster-Guided Unsupervised Domain Adaptation for Deep Speaker Embedding," IEEE Signal Processing Letters, vol. 30, pp. 643-647, May 2023.

d) Read the paper: W.W. Lin, M.W. Mak, N. Li, D. Su, and D. Yu, "A Framework for Adapting DNN Speaker Embedding Across Languages," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 28, pp. 2810-2822, 2020. Paper Code.

e) Read this lecture notes:
http://www.eie.polyu.edu.hk/%7Emwmak/notes/EIE558/EIE558_SpeakerRecognition.pptx
(Username: student Password: learning)

f) Watch these videos:

- "Advanced Pooling Methods for Robust Speaker Verification", *The Symposium on Speaker Recognition Research and Application (SRRW2022)*, Xinjiang, Nov. 2022. Video
- https://polyuit-my.sharepoint.com/:v:/g/personal/enmwmak_polyu_edu_hk/EQw5sYHayxNAha-_Cj90_G4BpVdnpGmft8rJ1-6-K5CA2Q?e=FVWWMw
- https://polyuit-my.sharepoint.com/:v:/g/personal/enmwmak_polyu_edu_hk/ESDIKJ2CDklHlQA7jV_UuU0BLCjj7czj0uuBe65OzjclPA?e=mXtZD9

5. Experiments for Learning Speech and Speaker Recognition

a) Repeat the experiments in Section 4(c) based on the code in https://github.com/Maohq97/Cluster-GuidedUDA

b) Incorporate MMD and consistency regularization in 4(d) to the program in 5(a)

6. References

1. Goodfellow, Ian, et al. Deep learning. Vol. 1. Cambridge: MIT press, 2016.
2. Y. LeCun, Y. Bengio and G.E. Hinton, "Deep Learning", Nature, vol. 521, pp. 436-444, May 2015.
3. M.W. Mak and J.T. Chien, Machine Learning for Speaker Recognition, Cambridge University Press, 2020.
4. Hansen, John HL, and Taufiq Hasan. "Speaker recognition by machines and humans: A tutorial review." IEEE Signal processing magazine 32.6 (2015): 74-99.
5. L.X. Li, M.W. Mak, and J.T. Chien, "Contrastive Adversarial Domain Adaptation Networks for Speaker Recognition", *IEEE Transactions on Neural Networks and Learning Systems*, 2021.
6. W.W. Lin, M.W. Mak, N. Li, D. Su, and D. Yu, "A Framework for Adapting DNN Speaker Embedding Across Languages," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 28, pp. 2810-2822, 2020. Paper Code.
7. Y.Z. Tu, M.W. Mak and J.T. Chien, "Variational Domain Adversarial Learning with Mutual Information Maximization for Speaker Verification," *IEEE/ACM Transactions on Audio, Speech and Language Processing*, vol. 28, pp. 2013-2024, June 2020.
8. Y.Z. Tu, M.W. Mak and J.T. Chien, "Variational Domain Adversarial Learning with Mutual Information Maximization for Speaker Verification, IEEE/ACM Transactions on Audio, Speech and Language Processing, 2020.

7 Computing Platform:

Amazon EC2, Google Cloud Platform, Microsoft Azure, Alibaba Cloud, Tencent Cloud, etc. (You have a budget of HK$2500)

- https://www.run.ai/guides/cloud-deep-learning/
- https://github.com/zszazi/Deep-learning-in-cloud
- https://cloud.google.com/gcp/getting-started

Appendix: Proposal Format

1. Introduction
   1.1 purpose of the project
   1.2 why is it important?

2. Background
   2.1 Background concepts and theories
   2.2 Existing methods in the literature
   2.3 Difference between your project and existing methods

3. Methodology
   3.1 System architecture
   3.2 Algorithmic details
   3.3 Software technology or methods for implementation

4. Preliminary Experiments and Results
   4.1 Pilot experiments (proof of concept)
   4.2 Preliminary results

5. Schedule

A good proposal should have the plan for the implementation. If the proposal has not implemented anything yet, the methodology part is usually very vague (which is not good). A good proposal should contain some concrete implementations.

In grant applications, preliminary results are very important, and they demonstrate that the proposers know what they are doing and will know the details on what should be done if they get the grant. But avoid giving the readers the impression that everything has been done; otherwise the proposers will not receive any grant. So, after the preliminary results, you should write something about what should be done next to make it a great project.