

Possible independent project ideas

CSCI-UA 9473

Augustin Cosse

January 2019

The point of this note is to provide a couple of suggestions (other ideas are of course conceivable) regarding possible independent projects for the class CSCI-UA 9473. Ideas marked with a star (*) will require more work. You are invited to work in groups of no more than 3 (ideally 2). Your results should then be summarized on a poster (e.g. 90cm x 122cm or 36W x 48H inches). [The posters presentation will be held around May 1st](#)). The poster should be [printed for April 25th](#).

1. Music and sound recognition

- Implement a music/sound recognition/classification algorithm (Find optimal latent representation). See for example the [GTZAN Genre Collection](#)
- Implement a simplified version of the Google assistant/simple chatbot (see for example the work of [Siraj Raval](#) or the work of [Tim Stein](#) for a start)
- Write a music composing algorithm (see for example [Magenta](#))
- Learn to recognize accents from speech recordings (see the [OSCAAR](#) webpage for data).

2. Computer Vision

- Learn how to process images for later use in autonomous driving systems (this would include segmentation, ROI detection, number extraction and digit/traffic lights recognition)
- Code a face recognition algorithm (see the [FERET database](#) or other databases on Kaggle)
- * Generate fake images (e.g. [fingerprints](#)) or fake videos with generative adversarial networks (GANs) (see the many examples from [youtube](#)).
- Learn emotions/facial expressions from images (see the [fer2013 dataset](#) or the [Cohn Kanade dataset](#))

- Training and Testing an autopilot (There exist several packages and resources that can be used to train cars, see for example [Donkey Car,medium](#)). There also exist many datasets that can be used to train autopilot algorithms (see for example [Udacity,NVidia](#), [Berkeley DD](#), [Dataset 3](#)). Also see [Yin and Berger, 2017](#), [Bojarski et al.](#)
- Learning depth maps or predicting steering angle from road images (see above for possible datasets)
- Object detection from satellite images (see for example the [Airbus challenge](#) on Kaggle)

3. Natural Language Processing (NLP)

- Extract information and develop a simple prediction algorithm based on Newspapers/twitter webpages
- Train a sentiment Analysis algorithm (see for example the [Amazon Customer reviews](#) dataset on Kaggle.)
- Develop simple Machine Translation algorithm ([data 1](#), [data 2](#))
- Detect personal attacks in online comments (see for example the [Wikipedia Comments Corpus](#))

4. Environment, Energy, Physics

- Predict Earthquake magnitude and location ([data](#), [source1](#))
- Detect salt bodies in subsurface imaging ([Data](#),[source1](#))
- Learn Physics (i.e equations) from Data

5. Economics/finance

- Implement a trading algorithm, Compare/benchmark various prediction strategies (see for example the [2 \$\sigma\$ challenge](#), or [Cryptocurrency data](#))
- Detect Fraud in mobile money transaction (see [PaySim data](#))
- Predict location and schedule in Vehicle/Bike Sharing ([NYC City bike dataset](#))

6. Social science

- Understand and investigate fairness and bias in ML based predictions (see for example [Beutel et al.](#), [Henzinger and Chen](#))
- Sentiment Analysis (see the NLP section).
- Improve movie recommendations on Netflix (see for example the [Movie-Lens DataSet on Kaggle](#))

7. Adversarial learning

- Study the effect of adversarial attacks/perturbations in the framework of face recognition/autonomous driving

8. Biology/ Medicine

- Predict gene/proteine expressions from various factors (e.g. histone modification signals)([Data](#),[Paper](#))
- Detect a medical condition from image data (see for example the website of the Special Interest Group on Knowledge Discovery and Data Mining (SIGKDD), [2006](#), [2008](#))

9. Reinforcement learning

- Train agents in simple games (e.g. [Pong](#), [Doom](#))

10. Computer Security/ Security

- Build a predictive model (classifier) capable of distinguishing between malevolent intrusions/attacks, and normal connections (see the [DARPA Intrusion Detection evaluation Dataset](#) the [KDD CUP 99 dataset](#),or the [ADFA Intrusion Detection Datasets](#))
- Improve airport security through early detection of threats (see the [Passenger Screening Algorithm Challenge](#)) (The prize might help you take a decision)