11-755/18-797
Machine Learning for Signal Processing
Final Poster Presentations

Instructor: Bhiksha Raj
TAs: Anurag Kumar, Abelino Jimenez, Yixuan Zhang

Fall 2017
1. **Vehicle classification based on magnetometer readings**

**using Machine Learning and Neural Networks**

Bob Iannucci, Cef Ramirez, Aggrey Muhebwa, Tong Lin, Yang Lu

We propose an implementation of a low-cost intelligent device bearing a three-axis magnetometer to sense, process, report and study traffic flow. The goal is to make this device capable of learning to automatically classify the type of passing vehicles and predicting their velocities. The challenges are; our sample size is small, and we need to use these magnetometer readings to create an optimized and efficient machine learning algorithm, capable of classifying car type and speed on a simple and low-power device. We test 8 different vehicles at 6 speeds, 48 cases in total. The trials are recorded by 3 three-axis magnetometers under traffic lane. Raw readings are noisy, we denoise the data by removing outliers using Hampel identifier and applying Gaussian smooth. We resample the data and transform it to frequency domain, we record the maximum magnitude of frequency in each 0.05-second window.

![Vehicle Classification Diagram](image)

We classify our data to 6 categories shown above and randomly select the readings of 36 cases for training and use 12 cases to test. We train an SVM (support vector machines) model and a Neural Network to perform classification. Using multi-class SVM with the assistance of Adaptive Boosting, we obtain a 100% accuracy in the training set. And we can correctly predict 9 out of 12 cases. On applying Neural Networks,

**Score this Project**

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
we get 100% accuracy on both training and testing data. We prove that Neural Networks can efficiently predict the vehicle type and speed in our data set. Further training and validation are needed when more data is available. Keywords: Machine Learning, Neural Networks, SVMs, Adaboost, Magnetometer Readings, Vehicle types, Velocities.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
2. **Music Composition with Long Short-Term Memory Networks**

Duo Cui, Zeqi Qiu, Yiqi Xiao, Haonan Zhou

Neural networks have been proven successful in many pattern recognition tasks. Recently, some researchers also tried to use neural networks to solve creative tasks like music composition. In this project, we implemented a Long Short-Term Memory (LSTM) Recurrent Neural Network to compose a melody with a given chord sequence input. Training data plays an important role in the quality of composed melody, so that a compositional goal should be clarified at first. We trained our network on chord/melody pairs which are extracted from pop music. The network learns which melodies fit to certain chord sequences by feed-forward and back propagation algorithm. Some of the generated melodies sounds pleasant. Others have noticeable differences from human compositions. In experiments, we trained different network topologies for the composition task, and evaluated the melodies generated by these networks by three metrics, including note range, number of notes played, and number of unique notes. The results show that as more layers are added to the network, the composed music drops in all three metrics, indicating a decreasing complexity.

**Score this Project**

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
3. Gender based Speech Emotion Detection

Xueting Hu, Qiunan Liu, Meng Li

As an important biometric feature like face expressions, human voice carries a lot of information relating to human emotions like fear, surprise, sadness, happiness. In this project, we will achieve this vocal emotion recognition by processing vocal signals and multi-classifying using machine learning techniques. Given a data set of 24 actors (12 males and 12 females) across ages speaking in 6 emotions, data augmentation is processed to get a handful training data. A bunch of feature extraction methods are applied, including RMSE, MAG (raw data), Spectrogram as well as MFCC. MFCC turns out to be the most efficient way to represent emotion information. Furthermore, data emotion evaluation and deduction are achieved by applying PCA and LDA. PCA with 45 eigen vectors gives the best result. As for the main classification algorithm, LDA, K-NN, SVM and GMM are applied. GMM (100 Gaussian distribution components) with a combination of MFCC feature extraction gives the best result for all 6 emotion classes, whose accuracy ranges from 79.0% to 96.0% with 87.5% average. 1-NN gives a second better result, while LDA and SVM do not performs well in this problem. In addition, a gender based pre-classification is implemented before emotion classification. Because gender itself carries lots of difference but not directly result in emotion detection. The average emotion detection accuracy increases by 37.6%.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
4. **Audio-based Visual Features for Large-Scale Audio Event Detection**

Ankit Shah, Anusha Kamath, Tejas Gokhale, Tyler Vuong

Convolutional Neural Networks (CNNs) have been used for numerous tasks such as classification, object detection and image processing. In this project, we explore the potential of CNNs to generate generalized features for audio. Specifically, we aim to visualize features learned within the CNNs for the problem of large scale audio-event detection using the AudioSet data. [1]. We train three CNN classifiers with each of the following inputs - MFCC, Log Mel Spectrogram, and Constant Q transform to investigate the features learned by the CNN at various layers. We visualize the feature hierarchy being modeled at each layer, analogous to what CNNs learn for images. We discovered that using Mel Spectrograms as input to the CNN gives a Mean Average Precision of 19.65 % an absolute improvement of 4 percentage points over other features training with just 20 epochs. We listen to the intermediate layer output post processing the CNN output using Auralization technique, and also visualize the deconvolved features. Our approach provides insights to better understand and deal with the pitfalls in network architecture and suggesting ways to improve performance.

**Score this Project**

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
5. Content-based movie recommendation system using emotions within movies.

Raj kiran Gupta, Ashwini Chall, Vaishnavi Ramesh

Content-Based Movie recommendation systems aim to recommend movies that users may be interested in. Although there exists a huge research on movie recommendations, the established movie recommendation techniques generally overlook various emotional aspects of movies. Content-based movie recommendations mostly rely on statistical correlations between movie ratings, genre, cast, etc., What user likes is the emotional connection with a movie, which is being ignored substantially by existing movie recommendation systems. Genre can depict only a very vague aspect of emotion. A movie will have more than one emotion at varying amounts like anger, love, sad, etc., Understanding the emotional graphs in a movie can get us closer to more accurate recommendation systems. There are several ways to understand how emotions changes throughout the movie. We can rely on either subtitles or speech(acoustic features) or visual features or all of them. In this project, we are exploring ways to use subtitles to derive emotion graphs of a movie.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
6. Implementation of Machine Learning Method for Pet Classification

Maisi Li, Minke Tang, Jianfeng Guo

This project is an attempt to apply traditional machine learning algorithm and neural networks to detect various kinds of cat aiming at seeing the performances of traditional ML methods and CNN, also seeing how far these methods can match the current industry standards in image classification. For this project, we mainly focused on K-nearest neighbor (KNN), Support Vector Machine (SVM), Back Propagation Neural Network (BPNN) as the representatives of traditional ML method, and three architectures of CNN.

The dataset we used for this project is from Oxford-IIIT Pet. The number of images in the dataset is 2000. In the detail, the dataset includes 10 kinds of cat, and each has 200 images. Since the similarity between some breeds of cat is high, it’s a challenge to find meaningful informative features that various from each cat breed. Also, we didn’t have enough training data due to the small dataset. Therefore, this project also focused on how to build an efficiency model for very little data to prevent overfitting in CNN training process.

First, we tried to extract 4 distinctive features (raw image pixel value, histogram, histogram of gradient and local binary pattern) and applied 3 classifiers (KNN, SVM, BPNN) on each feature. Secondly, we built 3 CNN networks from scratch. We tried different hyper parameters: layer number, filter size, filter number. Unsatisfied with the accuracy and time consuming, we used the bottleneck features of VGG 16 pre-trained network, trained the top layer and visualized the features.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
Based on our result, the first approach, traditional machine learning methods, did not provide the satisfied results. Even though these methods may be good at data mining, they are not the best choice to classify the complex dataset. The second approach, CNN, increased the accuracy. However, the training time and overfitting are two major problems we deal with in the CNN process. Finally, VGG pre-trained networks could efficiently train in just 10 mins to give us about 64% accuracy.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
Pyramidal cells are a type of nerve cell found in the cerebral cortex. These cells transmit information through electrical and chemical signals via specialized connections called synapses. These synapses are visualized using protein tagged fluorescent markers which binds selectively to these points. They appear as bright fluorescent spots or puncta and are rendered into 3D structures. Each structure is associated with morphological features which are used as data. The data was made available by Dr. Alison Barth, head of Barth Lab at the Department of biological sciences. The goal is to distinguish between the puncta that have different types of pre-synaptic associations using classification algorithms like Logistic Regression, SVMs and Random Forests. The logistic regression three-way classifier gave an accuracy of 65% on the test data. The location of the puncta was included as a feature and its impact was analyzed.
8. AUDIO FORENSICS FOR MARITIME RECOGNITION

Abhishek Kar, Yi-Ching Lee, Ramesh Oswal, Andrew Tracy

A maritime environment contains objects and activities that produce a unique set of audio signatures. A machine learning system trained on this unique set could be used to solve problems specific to this domain. This system could, e.g., classify the sound of an engine in the background of a phone call, which would allow listeners to cross reference known vehicles and roughly determine the caller’s location. The system could include many different maritime objects, such as animals, boats, or helicopters.

In this project, we have designed and built a system for automatic classification of marine objects in audio signals for use in forensic investigations. We have collected a dataset from sources on the internet and generated both hand-engineered and learned features, which we have tested with different classifiers. In this project, we limited the scope by using only two audio classes: boats and helicopters. The classifiers that we evaluated were gaussian mixture models and logistic regression.

The current results suggest more work is required in tuning the features and classifiers. However, because of the perceivable differences between audio files, we think that a highly effective classifier could be built using the framework that we have designed. We also believe this approach could be extended to a much larger audio learning system that could be used for forensic investigations in many different audio domains.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
9. CINEMATIC STABILIZATION

Matthew Baron, Samantha Wang

This project examines and implements a novel method for video stabilization based on cinematic camera paths. Traditional methods for stabilizing video using handheld devices filter the high frequency information from calculated camera paths to reduce jitter of video images. We use a sparse optimization approach to correct for unwanted frame motion by approximating each interframe difference in a recorded video as one of three types of linear motion. Our goal is to smooth out frame motion by estimating what the smooth camera path would be if the video was shot with more professional, mechanical methods such as a tripod for panning or a camera dolly for motion.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
10. **Adversarially Robust Face Recognition Algorithm based on Describable Facial Attributes**

Jihye Choi, Kiwan Maeng, Wonho Song, Sun Uk Kim

Machine learning is a technique that is profoundly changing our quality of life in various fields. The broad reliability on the machine learning also caused the increase of interest to understand the extent to which the machine learning algorithms are vulnerable. It has been shown that machine learning algorithms in general, and deep neural networks in particular, can be misled by attackers. In some cases, attackers can craft inconspicuous and physically realizable attacks to mislead machine learning algorithms. For example, in the application domain of face recognition, attackers can impose sophisticated pixel-level perturbations or craft eyeglasses which seem innocuous, and yet can be used to impersonate specific targets or be used to cause an arbitrary misclassification.

To tolerate these attacks, we propose a new attribute-based face recognition system that is more robust to the imperceptible attacks by its design. Instead of using traditional deep neural network structure to recognize faces, our system uses multiple VGG networks that output a probability of having a certain facial attribute, such as being ‘bald’, ‘young’, or ‘female’. Using the extracted attribute information, the fully-connected layers at the back-end recognize the faces. The system is motivated by the fact that human can recognize other’s faces even when there is perturbation in the image, by perceiving such attributes of the faces. And the image perturbations generated by previous attacks are focused on producing erroneous classification output, not the deep representations which are the extracted attribute information from VGG networks.

**Score this Project**

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
Our evaluation result shows that the proposed system not only show fairly successful recognition rate, but also are much more robust than the previous systems. Compared to VGG network and Openface, attacks applied to our system shows around 20% less success rate in succeeding attack, which is a quite noticeable percentile. More importantly, the attack that was generated by VGG network showed nearly 0% success rate in attack when applied to fool our system. By discussing the reason of such robustness, we also present insights into the nature of the vulnerability of a neural network.

**Score this Project**

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
11. CONTENT-BASED VIDEO RETRIEVAL OF SOCCER VIDEOS

Samuel Ishimwe, Yvonne Wambui, Tetevi Ekon

In our project, we address the problem of querying videos based on content matching in soccer matches. There is a lot of content generated by broadcasting companies and also soccer teams which now own television channels. There is a need to automatically summarize and analyse the videos for the viewers, to increase the fan experience. Our proposed system automatically partitions a video stream into video shots using color histogram to determine sequence of frames that are similar. They will form the video shot. Finding video shots of the same event is not an easy task because unlike images, context is very important. We scrap the websites which contain match comments text (gamecast) and we do language processing to extract information. We use the conditional random field model (CRFM) to detect the event shot boundary. In our experiments, we use an already broadcasted soccer match and perform automatic tagging of the different soccer events like goals, fouls and corner kicks. We are also able to identify the players involved in the events and the position of the field where the event occurred. All these form the tags that are the keywords used to query the video for specific events. The tagged shots can also be used as training datasets to develop machine learning algorithms that will identify soccer events.

Score this Project

Originality: 1-5 (5 is highest): 
Completeness: 1-5 (5 is highest): 
Overall: 1-5 (5 is highest):
12. GAUSSIAN MIXTURE MODEL FOR VOICE CONVERSION

Caleb Kaiji Lu, Tyler Nuanes, Serhan Oztekin, Nanshu Wang

Voice conversion is defined as modifying a source speaker's voice to sound as though it were produced by a target speaker. Such technology can have applications ranging from entertainment to speaker recognition. It would be beneficial to games and videos to reproduce the voices of famous actors or actresses, especially in translating media to different languages. It may also help in the grieving process (how many people wish they could hear a loved one's voice one last time after they pass?)

Voice conversion is a challenging goal, and a fully successful algorithm has so far proved elusive. Speech signals are created in a complex process involving vibrations of vocal chords and frequency filtering by the vocal tract. In addition, our auditory process is nonlinear, involving multiple frequency bands in the cochlea. Humans pick up characteristics of individual voices in frequencies ranging from 3 Hz to 10 kHz. Instead of attempting to develop a new algorithm, we focused our project on delving into a fundamental algorithm in the field and worked to reproduce it.

Our contribution is a complete implementation of the voice conversion workflow: feature extraction of the fundamental frequency $f_0$ and mel's cepstrum mcep; dynamic time wrapping (DTW) to align the speech of our target and source speaker in the training data; $F_0$ transformation to match the fundamental frequency to the target speaker; VQ and Full Conversion algorithms using Gaussian Mixture Models. Our training set includes 150 parallel audios from a female source and a male target. The results show that the MelCD, a measure of distortion, of Full Conversion is 37.66, and MelCD of VQ Conversion is 39.53. The Full Conversion reduced MelCD by 4.77 compared to the original distortion between source and target. Take this opportunity to listen to the audio signals yourself and hear what is possible!

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
13. A COMPARISON OF COMPRESSION ALGORITHMS FOR AERIAL IMAGES

Shreya Gupta and Aaron Perley

High resolution satellite imagery has many interesting applications, but the memory and network bandwidth required to store and transmit even small maps can be prohibitively expensive. However, aerial image tiles exhibit significant repetition of similar features, which suggests that large datasets, particularly over specific areas of interest, can be compressed significantly. This work presents a comparison of the compression performance on aerial images of two dimensionality reduction algorithms, convolutional neural network (CNN) autoencoders and principal component analysis (PCA). Although the CNN autoencoders demonstrated the expected tradeoff between compression ratio and reconstruction accuracy, the decompressed images exhibited significant blur, even at lower (5x) compression ratios. The PCA based approach learned unique basis images from the dataset well, but failed to generalize even across the training data. These results demonstrate that significant preprocessing of input images or specific modification of dimensionality reduction algorithms is necessary to provide acceptable performance on image compression.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
14. SPEECH NOISE SEPARATION THROUGH ICA METHOD

Mingyuan Yu

Image if you are talking to someone through cell phone in a noisy environment, it is desired to separate speech signal from noise mixing signal enable to improve the communication experience. The author design a program for doing this separation via Independent Component Analysis (ICA) method. The result shows if two signals meets ICA constraint, the output is outstanding great. However, if any one of the signal contain some time delay, which is highly possible happen in the real life, the output would sound poorly. Therefore, a technique of finding delay through cross correlation of two signals has been used.

Score this Project

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
In this project, we aim to analyze, recognize and classify human to human interaction through speech. The main and basic factor in each interaction is the emotion recognition. In general there are five main types of emotions, by means of which we can specify any emotional state of the speaker. These are called emotional dimensions: valence which contrasts pleasant against unpleasant emotions, activation indicates the level of arousal, power which characterizes whether the subject feels power and in control, expectation quantifies unpredictability of an event, and intensity which is a measure of perceived emotional intensity. We use an open source data set “SEMAINE”. It consists of audio recordings, emotional dimensions annotations and transcriptions from 150 participants. These sum for a total of 959 conversations of participants with operators, lasting approximately 5 minutes each. As the first step, we perform some pre-processings including resampling, pre-emphasis, and segmentation (frames). Then we extract speech-tailored features using 2-level hierarchical feature extraction method. At the first level, we extract 40 Low Level Descriptors (LLDs) including 12 Mel-Frequency Cepstral Coefficients (MFCCs), 26 filterbank energies (FBEs), central frequency (f_0), and Subharmonic-to-Harmonic Ratio (SHR). In addition, we calculate the delta regression coefficients of these LLDs with window size of W=2. This increase doubles the number of first level features (80). As the second hierarchical level, we apply 4 functionals, i.e. moments (standard deviation, skewness, kurtosis) and mean. In this case we have (4*80 = 320) feature dimensions. As the last step we train a Multi-input Multi-output neural network to classify all of five classes in parallel at each time sample. We look to explore and improve upon the abilities of predicting emotion efficiently through audio signal alone despite the inclusion of video and text.

**Score this Project**

Originality: 1-5 (5 is highest):

Completeness: 1-5 (5 is highest):

Overall: 1-5 (5 is highest):
16. KAGGLE COMPETITION: ZILLOW'S HOME VALUE PREDICTION (ZESTIMATE)

Pengxiang Hu, Vibhas Gejji

This project is an exploration and evaluation on the accuracy of Zillow's Zestimate prediction on residential housing price. In this report, we explored and analyzed the property information provided by Zillow. Based on the analysis, Boosted Tree and Artificial Neural Network algorithms are developed and applied to estimate the Zestimate accuracy.