Name(s)
Advait Arumugam

Project Number
S1501

Project Title
Cybian: A Machine Learning Program Based on Naive Bayes to Classify and Mitigate Cyberbullying

Abstract

Objectives/Goals
The objective was to create a program, called Cybian, to detect cyberbullying in social media content based on a set of algorithms to (i) classify the sender, (ii) evaluate and detect the bullying message by using machine intelligence based on Naïve Bayes and the category of the sender, (iii) send the classified bullying message to the administrator for review, and (iv) either send feedback to present the message to the recipient or take action.

Methods/Materials
I developed Cybian, a conceptual model based on machine intelligence that comprises of a set of algorithms to (i) intercept social media/instant messages at the recipient device, (ii) categorize the messages into pre-defined groups based on the sender, and (iii) classify messages by performing sentiment analysis, extracting features of bullying content, and comparing it against that of labeled data sets. Cybian determines behavioral patterns of the messages using the Bayesian probabilistic formula in the context of cyberbullying. (iv) Depending on the sender category and the threshold value, Cybian decides whether to present the message to the recipient or send it to the administrator. (v) Part of Cybian also resides on the administrator's device to present the classified message for review and respond by either sending feedback to the recipient or discarding the message with appropriate action. (vi) Based on the administrator's feedback, Cybian presents the message to the recipient. I developed a simulator to test my model using Anaconda Python 3.6.4 and Spyder 3. Separate modules for (i) the sender, (ii) the recipient, and (iii) the administrator were developed. To check the accuracy of the Naïve Bayes Algorithm, I created confusion matrices to compare classification and misclassification in the form of True Positives, True Negatives, False Positives, and False Negatives by performing tests on five different test data sets.

Results
Data from the five confusion matrices showed that Cybian had a 77.4% accuracy in detecting bullying messages.

Conclusions/Discussion
Cybian introduces a unique concept of categorizing the social media/instant messages based on the sender, detecting bullying messages by machine learning techniques, sending bullying messages to an administrator to allow or reject the messages. The results show that Cybian effectively mitigates cyberbullying while maintaining the privacy of the recipient so all my design criteria were met.

Summary Statement
Cybian, a hybrid software based on machine learning and human intervention successfully mitigates and ends the cycle of cyberbullying.

Help Received
I designed and programmed the algorithm myself after an internet search on techniques. I complied many databases together for this project.

Name(s)
Ramya Ayyagari

Project Number
S1502

Project Title
A Deep Learning Approach for Analyzing Tumor Histopathological Image Slides to Predict Breast Cancer

Abstract

Objectives/Goals
The goal of this project was to create a highly-accurate, automated detection platform that would be able to predict breast cancer by classifying histology slides from biopsies. The method must group images into benign and malignant subtypes, and should be integrated on a cloud server to eventually be easily accessible to any pathologist. This would greatly lower false positive/negative rates in diagnosis.

Methods/Materials
Using Keras and TensorFlow, I built two Convolutional Neural Networks (CNN), deep learning algorithms that were trained on the BACH (Breast Cancer Histology) dataset, to classify histology slides into binary (carcinoma, non-carcinoma) and multiple (normal, benign, in-situ, invasive) categories. The CNNs were then fine-tuned and run various times on an Amazon (AWS) GPU instance to ensure reproducibility of results. To achieve better performance, the CNNs were trained on top of deep, pre-trained neural networks, such as VGG16 and InceptionResNetV2. The CNN algorithms were then integrated into an automated detection platform that I built on the cloud.

Results
The fine-tuned binary classification CNN, when coupled with the pre-trained layers of the model InceptionResNetV2, achieved an accuracy of 95.0%. This is a significant improvement on the state-of-the-art accuracy for a similar dataset, 83.3%, indicating that my model surpassed a current gold standard. The multi-class classification CNN with VGG16 pre-trained layers achieved an accuracy of 83.75%, surpassing the state-of-the-art accuracy for a similar dataset, 77.8%.

Conclusions/Discussion
The contributions of this project include providing a clinically applicable platform to accurately diagnose breast cancer via tumor slides. This is important because the tedious work of classifying these slides manually is time-consuming, inaccurate, and costly. In the future, pathologists can use this to quickly validate their diagnoses, and patients will be able to receive results within seconds. Additionally, this project provides a deep learning framework that can be applied to other medical computer vision problems.

Summary Statement
I built a deep learning platform that can predict breast cancer with a 95% accuracy quickly and efficiently on a cloud server.

Help Received
I used the publicly available BACH dataset to train my model, and the open source platforms Keras and TensorFlow. I designed and performed all experiments myself.

Name(s)
Harris Beg; William Sun

Project Number
S1503

Project Title
skinCAM: An Effective Solution towards Skin Disease Diagnosis via Novel Deep Learning Algorithm

Abstract

Objectives/Goals
This project resolves a medical issue that many people face regarding dermatology. Skin diseases are common yet typically benign, but for an accurate diagnosis, patients have to visit a dermatologist, a costly endeavor. However, with the building of a convolutional neural network in this novel application of artificial intelligence, patients will be able to self-diagnose diseases that they may have.

Methods/Materials
Our neural network was created to train on a database of over 5000 distinct images. The chosen training sets came from trusted online databases that were validated by professional dermatologists (AAD, dermWeb, webMD). We built the model by creating five convolutional layers: each layer containing the convolutional filter for highlighting the features of the input, the ReLU layer as an activation function, and the pooling layer to minimize computation time. Then, we passed the output of the convolutional layers into a fully connected layer which used backpropagation and gradient descent to move the model towards optimal accuracy by updating the filters. The final output of the program was a probability value between 0 and 1, indicating the disease to the app.

Results
After training the neural network through 1000 iterations with a training step of 0.01, the prediction of the model produced an accuracy of 95.2% from 1000 tests. Each test ran the image through a separate distributable code that utilized the training data to confirm a valid prediction. Based off of the large testing sample size, the program can be statistically accepted since its margin of error is well within the pre-established design criteria.

Conclusions/Discussion
Our study demonstrates that an artificial intelligence approach can be utilized for an initial diagnosis, providing users with useful information about whether or not they should consult a dermatologist for further treatment. The app not only saves time and cuts expenses, but it is also equally as important by providing a unique tool for dermatologists around the world.

Summary Statement
We created a convolutional neural network to accurately diagnose skin diseases while also implementing the algorithm into an app for patients to use.

Help Received
The project was created entirely without outside help. However, we have recently reached out to Dr. Sicun Gao and Dr. Paravar for future studies in improving the accuracy of the model with powerful cloud-computing tools and with dermatologists through testing the app. We have also gotten help from

Name(s)
Guadalupe Bernal

Project Number
S1504

Project Title
Deep Learning Based Collision Avoidance Algorithm for Mobile Robots in Pedestrian Environments

Abstract

Objectives/Goals
The goal of this project was to create an algorithm that could effectively maneuver an autonomous robot by analyzing the video stream from a camera. The camera, mounted on a 2-wheeled robot, would send the video to a laptop for processing and determination of the next location to move to. The robot should be able to autonomously drive alongside people without collisions. The algorithm, with no prior knowledge of its environment, should guide a robot to follow objects without human intervention.

Methods/Materials
The system consists of a 2-wheeled differential robot using DC brushless motors and an Arduino Mega board, with a camera and laptop mounted on top. The camera sends the input video to the laptop which then feeds the frames to a neural network which recognizes the objects. The location and class of the object is then sent to an algorithm which analyses their 2D position and sends commands back to the Arduino via a bluetooth module to control the motors. The algorithm was developed in C++ using the computer vision library OpenCV 3.1 and consists of a processing pipeline with an initialization stage and a single processing loop. The first step is the frame acquisition which captures the frame, then a modified YOLO neural network locates the objects in the frame. The distance and angle to the objects is found using their positioning on the screen, and finally the next location is calculated and the speed of the motors is transferred to the microcontroller. I designed the robot in SolidWorks, welded the frame as well as 3D printed parts with ABS filament.

Results
I tested the robot with myself and a variety of different objects detected by the YOLO neural network in both indoor and outdoor settings. The robot managed to follow balls and other toys while avoiding people in the way. The program runs at approximately 3 - 3.5 fps making it functional for real-time usage, and has a 1 - 5% error when calculating distances between 1 and 5 meters away.

Conclusions/Discussion
At its current state this robot serves as a successful testing platform for research and investigation purposes. The algorithm is gateway to show the possibilities that exist for the transferring of advanced recognition softwares from high-end to low-end technologies, like laptops and phones. This robot can currently successfully detect an object, find its locations, and follow it while avoiding collisions with people.

Summary Statement
I designed and built a robot, modified a neural network that detects the objects within a frame, and wrote an algorithm to locate the 3D position of the objects. The robot can navigate by itself while avoiding collisions.

Help Received
I took engineering courses at my high school that allowed me to learn programming and 3D computer modeling. I have also been a part of multiple robotics teams including the FTC competition where I was the main developer. My main source of information came from the internet.

CALIFORNIA SCIENCE & ENGINEERING FAIR
2018 PROJECT SUMMARY

Name(s)
Makenzie A. Dorsey

Project Number
S1505

Project Title
Mathematical Origami

Abstract

Objectives/Goals
My objective was to determine how origami can be used to solve for the roots of a cubic polynomial. The goal was to discover an easier method of solving for polynomials.

Methods/Materials
The materials consisted of graph paper, a ruler, and a pencil. I utilized a method created by Austrian engineer, Eduard Lill, that involved drawing the polynomial on graph paper, making a fold using two points on the graph, and finally, using trigonometry to solve for the roots of the polynomial.

Results
The accuracy rate of solving for the roots of a cubic polynomial was the highest among polynomials that had solutions with multiplicity. Polynomials with real, rational solutions were the seconds highest while polynomials with imaginary solutions had the lowest accuracy rates. Throughout all of the data collected, the average accuracy rate was above 90%.

Conclusions/Discussion
The process of solving for the roots of a cubic polynomial with origami is indeed possible. Origami may not be able to solve for imaginary solutions but it can be solve for both rational and irrational solutions. This means that there is an alternate way of solving for the roots of a polynomials without using the Rational Zeroes Theorem.

Summary Statement
I measured the accuracy rate of solving cubic polynomials using origami.

Help Received
One of the math teachers at my school provided guidance and assistane when I was organizing my project.

Name(s)
Dhanvee Ivaturi; Philip Kabranov

Project Number
S1506

Project Title
Improving Breast Cancer Detection in Fine-Needle Aspirate Biopsies through Machine Learning

Abstract

Objectives/Goals
The intent of this project is to compare the accuracy multiple machine learning (ML) algorithms to improve and/or automate breast cancer detection. This research takes into account various ML algorithms (logistic regression, support vector machines, and neural networks) and PCA dimensionality reduction.

Methods/Materials
Python, along with the Jupyter Notebook development environment are used in this software. The SciKit Learn and Tensorflow libraries are imported to implement the machine learning algorithms. The pandas and NumPy libraries are used for data visualization. The dataset in this project is the Wisconsin Breast Cancer dataset, which contains the records of 569 patients as unique 30-dimensional feature vectors. These features are extracted from measurable visual attributes of tissue samples. The 569 entries are randomly split into test and training subsets, and then used to train and test a machine learning. The algorithms are also trained separately with a dataset processed through PCA in order to eliminate features with low impact on the outcome.

Results
The dataset was found to be linearly separable, meaning that a logistic regressor and support vector machine could be applied to this dataset. The various ML algorithms were all trained using the same data source. For both data processed by PCA (reduced to 5 features) and non-reduced data (30 features), the logistic regression algorithm produced optimal mean accuracy over 20 training cycles: 97.82% for the non-reduced 30-dimensional feature set, and 97.24% for the feature set reduced to 5 dimensions, while the SVM produced a lowest mean accuracy of 96.69% and 96.50%, respectively. These accuracies are comparable to or exceed those of medical professionals.

Conclusions/Discussion
The results demonstrate that the best algorithm for this dataset is the logistic regressor, being the fastest to train and the most accurate. Neural network accuracies are in between those of the logistic regressor and SVM but takes 100 times longer to train. This is likely due to the computational complexity of the backpropagation training algorithm for neural networks. Also, the relatively low performance of the SVM is likely due to the fact that the data is not perfectly separated into two non-overlapping clusters, giving way to classification errors.

Summary Statement
Applied and analyzed the accuracy of multiple machine learning algorithms and a neural network on a dataset consisting of numerical values that relate to image attributes from fine-needle aspirate biopsy samples.

Help Received
The implementation of the machine learning algorithms and results were reviewed by a machine learning professional.

CALIFORNIA SCIENCE & ENGINEERING FAIR
2018 PROJECT SUMMARY

Name(s)
Sohini Kar

Project Number
S1507

Project Title
Factorizing Delayed Powers of Generalized Fibonacci Sequence

Abstract

Objectives/Goals

Let a_{n} denote the number of positive integers containing only the digits 1, 3, and 4, so that the sum of the digits is equal to n. Then a_{2n} is always a perfect square, for any n. (In fact, it is the square of a Fibonacci number.) It appears that the natural recurrence relation generating the a_{n} s, factors through the recurrence for the Fibonacci numbers.

Hypothesis: For a second-order linear recurrence relation of the form u_{n} = (p)u_{n-1} + (q)u_{n-2}, we can create a recurrence relation which yields sequences which have an = u_{n}^(k) or n_{k}= u_{n}^(k). We can also create a closed-form generating function and a combinatorial/pictorial representation.

Methods/Materials

Project Materials: Computer with Matlab and Maple installed

General project methods: Generating functions, recurrence relations, partial fraction expansions, Hadamard products, tilings, Cauchy's residue theorem

Results

First, I found the generating function of u_n = u_{n-1} + u_{n-3} + u_{n-4}. Additionally, I used the generating functions for the square and cube of Fibonacci sequence and its generating function to derive a recurrence relation for the kth power of Fibonacci sequence.

Next, I investigated the kind of recurrence relations that yield square of various second order linear recurrence of the kind defined in the hypothesis, un = (p)u_{n-1} + (q)u_{n-2}. To do this, I especially used Hadamard products and Cauchy's residue theorem. Also, I used diagrams, especially tiles, to model the problem.

Then, I used the findings for a new recurrence relation that can be generalized so that the kth term is the kth power of the second order linear recurrence. I found a closed-form generating function for this, specifically finding the numerator, which no one has done before.

Finally, I explored different possibilities for combinatorially represent my solution, specifically a pictorial form.

Conclusions/Discussion

My project has strong implications for furthering the field of steganography, as the recurrence relation and generating function may be used to encrypt messages in the images better. It has potential to be used with Zeckendorf's Theorem, which states that every number can be represented using distinct Fibonacci numbers.

Summary Statement

I created my own number sequence, based on a pattern I had already observed, and derived a recurrence relation and generating function for this sequence to generalize it; I also found a way to pictorially represent my project.

Help Received

My mentor, Simon Rubinstein-Salzedo, offered me valuable guidance throughout this project.

Name(s)
John Kim

Project Number
S1508

Project Title
Emotion Recognition from Human Speech Using Temporal Information and Deep Learning

Abstract

Objectives/Goals
Emotion recognition from voice by machine is a challenging task, but it has great potential to make empathic human-machine communications possible. In conventional approaches that consist of feature extraction and classifier stages, extensive studies have been devoted to developing good feature representations, but relatively little effort was made to make proper use of the important temporal information in these features. The goal of this research is to develop a model combining features known to be useful for emotion recognition and deep neural networks to exploit temporal information when recognizing emotion status.

Methods/Materials
I propose a model that combines a feature extraction stage known to be useful for emotion recognition and a Deep Neural Network (DNN) to model the unknown mechanism in recognizing emotion status from the temporal sequence of feature vectors, where the DNN consists of Convolutional Neural Networks for local and global convolution and Long Short-Term Memory layers. Two different model structures are developed and compared to conventional approaches. Considering the practical use of emotion recognition systems, the performance gap issue between speaker-dependent/independent modes is addressed. The performance evaluation is performed on the Berlin Emotional Speech Database, which is one of the most widely used databases. The database consists of 535 utterances from 10 talkers, each utterance representing one of seven different emotions.

Results
A benchmark evaluation demonstrates that the proposed model achieves 88.9% recognition rate, replacing the state-of-the-art performance of 86% with a big margin equivalent to 20.7% error reduction rate. A deeper analysis validates that the emotion space formed by internal representations of the proposed model is similar to that of human perception. The recognition rate is degraded by 7%, when the model was evaluated in speaker-independent mode.

Conclusions/Discussion
A novel model is proposed to recognize emotion from human speech. The model consists of acoustic speech analysis to maximize the efficiency of well-known feature extraction and DNN to learn an unknown mechanism of temporal information. The proposed model replaces the state-of-the-art performance with a big margin. Further works include using an extended feature set, evaluation on different databases, and performance improvement in speaker-independent mode.

Summary Statement
A novel DNN model is proposed to recognize emotion from human speech, and achieves 88.9% recognition rate, replacing the state-of-the-art performance of 86% with a big margin equivalent to 20.7% error reduction rate.

Help Received
Dr. Rif Saurous guided me on my research, specifically on the problems of cross-validation.

CALIFORNIA SCIENCE & ENGINEERING FAIR
2018 PROJECT SUMMARY

Name(s)
Elizabeth Kravtchenko

Project Number
S1509

Project Title
Who Are the Winners? Balancing the California State Science Fair Award System

Abstract

Objectives/Goals
The purpose of this project was to analyze thousands of publicly available historical records and understand (a) who the winners of the California State Science Fair are and (b) how the school grade affects student's probability of winning a gold medal. My hypothesis was: if a student's grade increases, then the student's probability of winning gold also increases.

Methods/Materials
In my research and analysis, I analyzed the last 10 California State Science Fairs. My independent variable was school grade. My dependent variable was the probability of winning a gold medal. In my modeling and simulation, I developed a mathematical model to balance the award system, and ran a computer simulation to show how the new model would even out the probabilities of winning gold in the senior division. I used the following materials: historical data from cssf.usc.edu, PC, and Microsoft Excel.

Results
My analysis showed that in each division, as a student's grade increased, the probability of winning gold also increased. However, I discovered the discontinuity between the two divisions, which was caused by the "aging up" effect, similar to the one you witness in sports. When students "aged up" from the junior division to the senior division, their probability of winning gold significantly decreased. This probability rebounded and exponentially increased as students advanced to higher grades.

Conclusions/Discussion
My hypothesis was correct! I developed a mathematical model to address the "aging up" discontinuity. I proposed to split the senior division into two divisions (grades 9-10 and 11-12) and double the number of students in these new divisions. I ran a computer simulation to show how the suggested award model would even out the probabilities of winning gold among high school students. This innovative model would make it more attractive and motivating for freshmen and sophomores to continue participating in science fair competitions.

Summary Statement
I developed a mathematical model to balance the California State Science Fair award system and ran a computer simulation to show how the new model would even out the probabilities of winning gold in the senior (high school) division.

Help Received
David Talcott (physics teacher), Michael Skrable (math teacher), Addison Lewellen (English teacher), and Bob Dubrow (mentor). They reviewed my project and presentation, and provided feedback and suggestions. Vladimir Kravtchenko (my Dad) taught me how to extract and clean data and to use

CALIFORNIA SCIENCE & ENGINEERING FAIR
2018 PROJECT SUMMARY

Name(s)
Rachana Madhukara

Project Number
S1510

Project Title
Asymptotics of Character Sums

Abstract

Objectives/Goals
In this project, we aim to prove certain properties about a particular function c(n) = b_nr(n). This is where b_n is a Boolean function with b_n being 1 if n = x^2 + y^2 for some integers x and y or 0 otherwise and r_chi(n) is the sum of all of the Dirichlet characters of d, where d divides n. The function c(n) sums the all of the chi values of the divisors of a certain number n if and only if n can be expressed as the sum of two squares. Therefore, the question we ask is the following: What are the asymptotics of the character sums of the function c(n)?

Results
In order to investigate this problem, we first represent the character sum of r(n) as an asymptotic and prove that the asymptotic is roughly L(1, chi) with a small error term. Additionally, we compute a representation for the character sum c(n) as an Euler product, and also find error bounds on the asymptotic for the character sum.

Conclusions/Discussion
We analyzed the asymptotic, or growth rate, of a very special function c(n) which describes a very particular group of primes. In specific, our growth rate describes the group of primes which are dependent on two character values. Additionally, we found some error bounds on how accurate our asymptotic is.

Summary Statement
In this project, an asymptotic for a function c(n) was found along with an error term using elementary number theory techniques.

Help Received
My mentor Dr. Simon Rubinstein-Salzedo greatly helped me with this project. His main role in the project was always leading me in the right direction. He did this by providing me with lots of relevant papers to read and giving me ample suggestions.

Name(s)
Anish R. Neervannan

Project Number
S1511

Project Title
A State-of-the-Art Approach to Detect Diabetic Retinopathy Using Convolutional Neural Networks and Adversarial Networks

Abstract

Objectives/Goals
The purpose of this project was to determine if a state-of-the-art generative adversarial network (GAN) could be used in conjunction with a convolutional neural network (CNN) image recognition algorithm to detect diabetic retinopathy in fudus photography more accurately than an independent CNN and match the accuracy of a medical professional's mental model.

Methods/Materials
The materials used included a System76 machine; a dataset of 35,000 retinal scans downloaded from Kaggle's diabetic retinopathy homepage (available for public use); the Inception Resnet v2 model; and Tensorflow, Keras, and other machine learning libraries. The image dataset was downloaded and retinal scans were randomly split into training & validation sets and test sets. Multiple programs were written in Python to crop, reshape, visualize, and transpose the scans. Keras and Tensorflow libraries were used to build the Convolutional Neural Network architecture on top of the Inception base model and train the algorithm. A separate Generative Adversarial Network architecture was built, the training data was fed through, and the generated images were added back to the original dataset for another round of training. 200 images from the test set were passed into both trained models and sent to three medical professionals. These predictions were used to compute precision, recall, and accuracy.

Results
The GAN+CNN algorithm's accuracy was 72.00%, the CNN-only algorithm's accuracy was 74.00%, and the three medical professionals' accuracies were 76.00%, 84.00%, and 88.00%. On scans without diabetic retinopathy, the CNN-only algorithm got a precision score of 75.00% and a recall score of 100.00%, the GAN+CNN algorithm got a precision score of 87.80% and a recall score of 100.00%, and the medical professionals got precision scores of 97.15%, 92.69%, and 88.50% and recall scores of 70.00%, 29.39%, and 71.24%.

Conclusions/Discussion
The hypothesis was not supported; even though both computer algorithms came close to the medical professionals' accuracies, it did not match them. In addition, all tests indicated that the GAN + CNN combination classifier came close to did not outperform the CNN-only classifier. However, this report broke new grounds in that the CNN-only classifier achieved a slightly higher accuracy than that of last year's project, which had outperformed a prior model that was published in a Stanford paper by an 8% margin.

Summary Statement
I developed a generative adversarial network to work in conjunction with a convolutional neural network to detect diabetic retinopathy nearly as accurately as a trained medical professional.

Help Received
Dr. Kapil Kapoor, Dr. David Chia, and Dr. Sanjay Kedhar were the three ophthalmologists that classified my test set data for comparison with my algorithms' results. My parents, Raj and Aruna Neervannan, and my science teachers, Mr. David Knight, Mr. Tim Smay, and Mr. Nicholas Brighton, reviewed my report

CALIFORNIA SCIENCE & ENGINEERING FAIR
2018 PROJECT SUMMARY

Name(s)
Jian Park

Project Number
S1512

Project Title
On the Modular Properties of Hypothetical Collatz Loops

Abstract

Objectives/Goals
The purpose of this project was to observe inherent modular trends within all hypothetical Collatz loops.

Methods/Materials
The Collatz function is defined as follows: start with some positive integer x. If x is odd, multiply it by three and add one, and if x is even, divide it by two. A recursion of this function creates a Collatz series, and the Collatz conjecture predicts that no matter what initial value x is chosen, any Collatz sequence will eventually reach 1. One possible scenario that would disprove this conjecture is if there existed a loop that did not include the number 1. In order to explore these hypothetical loops, I first developed the Collatz modulo web, which is a method that can be used to compute the possible modulo values of the elements in a general Collatz sequence. In order for a loop to be real, it must also exist within the modulo webs, so I developed a depth-first search algorithm that traversed all possible trajectories within the modulo webs to find loops with length n.

Results
By using the traversing program I developed, I was able to test up to n=24, which yielded no non-trivial loops.

Conclusions/Discussion
The Collatz modulo web concept can predict the modulo values of Collatz sequence numbers, and combined with the searching algorithm, it can be used to computationally calculate the existence of loops by loop length, rather than by initial value. This new method to check for loops may bring new insight into the previously unsolved problem.

Summary Statement
I applied modular arithmetic to a generalized formula in order to develop the Collatz Modulo Web, which is a concept that can help identify potential Collatz Loops.

Help Received
I derived all of the mathematical concepts myself. My father assisted me during the develop of the program I used in the second portion of my project.

CALIFORNIA SCIENCE & ENGINEERING FAIR
2018 PROJECT SUMMARY

Name(s)
Leonardo S. Park

Project Number
S1513

Project Title
The Banach-Tarski Paradox and Equidecomposability and Their Applicability to the Infinite

Abstract

Objectives/Goals
The objective of this project was to formulate two mathematical proofs(one through a graphical demonstration in the context of the mechanisms of the Hyperwebster, an infinite dictionary, and one through compiling) to affirm the validity of the Banach-Tarski Paradox.

Methods/Materials
Personal computer, Painting Software(Paint), and the Pyth Compiler/Executor.
This project was carried out through two experiments to show how it is possible to create two duplicates from one starting sphere. The first experiment was through the computer and the software, Paint, to draw 8 graphs and sketch lines to represent a hypothetical three-dimensional sphere rotating in a graphical plane and making its mechanisms imitate that of the Hyperwebster, an infinitely countable dictionary. The second experiment was conducted through the software Pyth Compiler, to basically code a three-dimensional sphere and also "rotate" it to highlight how its movement is representing the workings of the Banach-Tarski Paradox.

Results
I was able to prove that the Banach-Tarski Paradox was valid by converting the mechanisms of the Hyperwebster, an infinitely countable dictionary, to a three-dimensional graphical sense for a solid sphere and by utilizing the Pyth Compiler/Executor to compile a rotating sphere to imitate the workings of the Banach-Tarski Paradox.

Conclusions/Discussion
I was able to confirm my hypothesis of the legitimacy of the Banach-Tarski Paradox in a pure mathematical sense. I have realized that currently, it is only able to be proven in the abstract mathematical world because in this context, this paradox is not constrained by the limitations of the physical boundaries that are present in the real world. However, it is quite true that what is able to be proven mathematically is often proved through physics as well, and it may be possible for an application in the physical world to be created if our technology ever reaches that level.

Summary Statement
I devised two mathematical proofs through a graphical context and through a compiler to test the validity of the Banach-Tarski Paradox, which states that one 3-D ball, in set-theoretic geometry, can be cut out to yield two identical copies.

Help Received
I did not receive any assistance and worked alone on my experiments.

Name(s)
Rami Ratl Mrad; Carey Yoon

Project Number
S1514

Project Title
Music Measure Theory: A Mathematical Analysis of Musical Dissonance and Consonance

Abstract

Objectives/Goals
The objective of our project is to find the accuracy and reliability of the Music Measure Theory.

Methods/Materials
Upright Piano, Computer w/ internet access, Sheet music (Johann Sebastian Bach's Prelude in C Major BWV 846), Calculator, Paper and pencil: Found all frequencies of all keys on piano using computer, used intervals from the sheet music to generate ratios and graphs, calculated accuracy.

Results
Consonant-Consonant: 100% Accuracy
Consonant-Neutral: 100% Accuracy
Consonance-Dissonance: 100% Accuracy
Dissonance-Neutrality: 50% Accuracy

Conclusions/Discussion
The Music Measure Theory makes mathematical sense and is accurate, so now people know why sounds sound consonant or dissonant together. In the past, there was no solid proof of the theory. People can use this theory to expand their current knowledge of sound by building off of it.

Summary Statement
We mathematically analyzed sound frequencies with the Music Measure Theory in order to prove or disprove its accuracy.

Help Received
None. We designed and performed the experiments ourselves.

Name(s)
Alejandro Rodriguez

Project Number
S1515

Project Title
BACON: Deep-Learning Powered AI for Poetry Generation with Author Linguistic Style Transfer

Abstract

Objectives/Goals
Can machines create art? Can a computer write meaningful poetry? The goal of this project is to develop a software system that automatically generates poetry (in English) with the style of any given author, what is known as linguistic style transfer.

Methods/Materials
The problem stated above is split into the following 3 sub-problems:
(1) Linguistic style modeling: this is achieved by extracting, from the author's text corpus, the following linguistic features: (a) high-entropy n-grams: by building a TF-IDF (Term Frequency - Inverse Document Frequency) model; (b) relevant themes and topic words: by building an LDA (Latent Dirichlet Allocation) model
(2) Style transfer: use these probabilistic models to guide the automatic generation of poems. Achieved by probabilistic boosting of the previous features in the language model used by an Automatic Poem Generator (APG)
(3) Automatic poem generation: generate meaningful poetry (content) with rich aesthetic rules (form). This is built as a pipeline of a LSTM (Long Short-Term Memory) RNN (Recurrent Neural Network) -which addresses poem content generation-, and a WFST (Weighted Finite State Transducer) -which addresses poem form shaping-.

BACON, a basic software prototype that addresses the above mentioned methods, has been developed. It consists of several Python and Lua scripts, on a software platform built with several open source ML/DL modules and that extends an APG implementation developed by D. Kiela (Facebook AI Research) and J. Hopkins (Univ. Cambridge, UK).

Results
The quality and human-likeness of the poetry produced by BACON was tested through an extrinsic evaluation procedure, involving 62 human participants. The results indicate that participants were unable to tell the difference between human and computer-generated poems in any statistically significant way (Chi-squared test, p=0.85).

Conclusions/Discussion
The poetry generated by BACON approaches human poetry in aesthetic quality, while capturing some linguistic features of an author's style. Therefore, the answer to the question posed in the project goal is positive. Future work will expand the set of features subject to linguistic style transfer.

Summary Statement
I developed a deep-learning powered AI software system that automatically generates poetry in the style of any given author.

Help Received
I initially reached out to D. Kiela (Facebook AI Research) and J. Hopkins (Univ. Cambridge, UK) for guidance about feasibility and about possible approaches. After that, the solution workflow, methods, and programming code that are described in the project were developed by me.

CALIFORNIA SCIENCE & ENGINEERING FAIR
2018 PROJECT SUMMARY

Name(s)
Grant Sheen

Project Number
S1516

Project Title
Wireless Brainwave Classification for Alzheimer's Patients via Efficient Neural Network Computation

Abstract

Objectives/Goals
In order to break the communication barrier between Alzheimer's patients and their caretakers, there is an inevitable need to create a thought recognition software to classify the brainwaves of Alzheimer's patients recorded by a portable and wearable wireless device. The unique characteristics of wireless brainwave data (limited amount and low resolution) pose a significant challenge to standard neural networks, which require abundant data. I developed an efficient neural network model and a new training algorithm to attain high classification accuracies of wireless brainwave data, which standard neural networks fail to do.

Methods/Materials
The brainwaves were recorded with a 14-electrode wireless headset called EPOC+ by Emotiv (the data were obtained from the UC Irvine MIND Institute and were de-identified). To overcome the challenge of limited data, I constructed a dimensionally reduced neural network model trained by my Alternating Minimization (AM) algorithm. During training, I computed the unknown parameters of the neural network by minimizing a non-smooth and non-convex objective (cross entropy) function one variable at a time while fixing the rest, until convergence of the objective values. I discovered that the objective function in each variable is piecewise convex, so the global minimum can be computed with bisection. The overall iterative AM algorithm is descending and convergent, free of the step size (learning) parameter in the standard gradient descent method. After training, the testing data is fed into the neural network and the class probabilities are calculated for prediction of thoughts.

Results
After 7 months of research and development, my algorithm classified 4 daily thoughts of an Alzheimer's patient at a 90% accuracy. This indicates that my software can be used to restore the communication capabilities of Alzheimer's patients.

Conclusions/Discussion
My neural network model provides an effective solution to the communication issue between Alzheimer's patients and their caretakers. My algorithm works with a wireless headset, allowing thought recognition to be run anywhere in real time. I was the first person to design a dimensionally reduced neural network for brainwave data, develop an Alternating Minimization algorithm for neural network training, and derive analytical formulas for curved decision boundaries.

Summary Statement
I developed an efficient neural network model and a new training algorithm for wireless brainwave classification of Alzheimer's patients.

Help Received
I designed the neural network model and developed a new training algorithm myself. I received help from Professor Knut Solna of the Mathematics Department at UC Irvine in understanding background material.

CALIFORNIA SCIENCE & ENGINEERING FAIR
2018 PROJECT SUMMARY

Name(s)
Patrick C. Song

Project Number
S1517

Project Title
Early Classification of Alzheimer's Disease Using Machine Learning

Abstract

Objectives/Goals
Alzheimer's Disease (AD) is the sixth leading cause of death in the United States. Despite its significance, there is currently no cure for the disease. The early diagnosis of AD is essential for patient care and relevant researches. The purpose of this project is to create a machine learning classifier that can accurately distinguish between healthy subjects and subjects with AD. This classifier can potentially aid doctors in diagnosis and treatment.

Methods/Materials
Data and MRI images from the Alzheimer's Disease Neuroimaging Initiative were used to train the classifier. Using SPM 12, selected features were extracted from the MRI images. These features were input into the neural network, which was coded using MATLAB. The overall accuracy, test performance, and sensitivity of the network were calculated in order to grade the classifier.

Results
The artificial neural network learned to classify between subjects with Alzheimer's Disease and normal subjects at a 95.6% accuracy level. Additionally, volume of cerebrospinal fluid, Mini Mental State Examination, and volume of gray matter were found to be the most influential features on the network.

Conclusions/Discussion
In this project, I created an artificial neural network that outperformed several published Alzheimer's Disease classification approaches. Though far from being able to replace a neurological radiologist, the classifier is of great benefit in prioritizing scans for radiologists to analyze and identifying important biomarkers for the disease.

Summary Statement
I created an artificial neural network that can accurately distinguish between subjects with Alzheimer's Disease and healthy subjects, and identified the most influential features on the network.

Help Received
I developed the code for my project on my own. My high school science teacher, Mr. David Van Muyden, proofread the project and gave hepful guidance.

Name(s)
Ashwin Viswesvaran

Project Number
S1518

Project Title
Analyzing 2D-Convolutional Neural Network's Ability to Learn Obstacle Avoidance in Self-Driving Applications

Abstract

Objectives/Goals
Analyze the ability of a 2D Convolutional Neural Network to learn random 3D objects as obstacles by training it with fixed set of 3D objects of varying color and orientation placed randomly on the track.

Methods/Materials
Added a Hall Effect sensor and related code changes to the "Donkey" open source based self-driving car project which lacked consistent speed control and precise maneuverability so that it could be used with my project. After mastering repeatable autonomous driving around a test track loop, several iterations of training placing a pre-selected list of various 3D objects and avoiding them was carried out. Training included systematically varying the independent variables such as color of the given object, its orientation and its position on the track. After training, a random set of new 3D objects where used to evaluate the learning achieved in general and to see how each independent variable impacted the effectiveness of learning in a quantitative manner. Experiments were carried out in controlled environment (light condition, speed, track etc.) to ensure the results observed could be repeated.

Results
When 3D objects used during training and random new 3D objects not seen before by the car were randomly placed at different orientations and at random places on the track during testing, the car avoided them as obstacles in most conditions, failed to maneuver effectively or failed to detect the object as obstacles in others.

Conclusions/Discussion
The experiment clearly demonstrated that 2D CNN with a single camera input could be used to achieve obstacle avoidance. However, the effectiveness of the results depended upon specific aspects of the independent variables. For example, darker objects were avoided much predictably than lighter colored ones. Obstacles were avoided consistently on straight sections than on curves. Both quality and quantity of training had a great impact on the observed behavior of a run.

Summary Statement
The obstacle avoidance behavior observed during testing with random new 3D objects which differed significantly from the training objects clearly demonstrated 2D CNN's ability to learn obstacle avoidance.

Help Received
I built the car per Donkey project and added Hall Sensor. Discussed new concepts with father. My mentor helped point to useful research. Parents helped placing objects on track as I trained.

Name(s)
Mokshith Voodarla

Project Number
S1519

Project Title
TionAI: Understanding Human Emotion through an Ensemble of Convolutional Neural Networks for better AI-Human Interaction

Abstract

Objectives/Goals
To create a model that can understand human emotion in such a way that it essentially "feels" emotion just like humans do. Emotion is a key enabler of natural human interaction so AI understanding this idea would not only enable more proactive computers which can interact with humans just like other humans do, but it would also enable AI to better diagnose mental health issues like depression through this same mean of interaction.

Methods/Materials
Devised an ensemble (an interconnected group) of 3 convolutional neural networks (CNNs) to understand various aspects of emotion in different visual settings (images). One understands emotion in backgrounds/setting (EnvoNet), another understands emotion in objects/foreground (SubjectNet), and the last one synthesizes the decisions from the other 2 CNNs by assigning weightage to each CNN in the final decision made (DeciderNet).
Utilized the Wheel of Emotion, an idea that contrasts and characterizes emotions, to derive emotion categories for the ensemble to understand.
Wrote a Google Image query script to recursively search through trees of search suggestions to compile datasets for each CNN separated into emotion categories.
Trained CNNs through transfer learning, calculating bottleneck values through the penultimate layer of a given pretrained model to gain insights on how to change network weights and biases.
Designed a method for testing optimal loss functions and optimizers to yield maximal training results.
All datasets split into train, validation, and test (80-10-10) to compute true ensemble accuracy.
All training was done using Python 2.7 and TensorFlow, Keras, and other stock math libraries on an NVIDIA GPU.

Results
Created a network ensemble with validation accuracy of 90.2% overall, exceeding expectations of standard CNNs in such a task. The AI started understanding ideas it wasn't taught through training as well (i.e. associating knives, guns, etc. to various negative emotions).

Conclusions/Discussion
I trained an ensemble of CNNs to gain visual understanding of human emotion, through transfer learning and various optimization techniques. Performing this approach over an ensemble proved to be effective, and improvements to data collection and bias can further improve ensemble accuracy. The AI is almost indistinguishable from humans in understanding emotion in various images.

Summary Statement
I created a model which computers can use to "feel" emotion just like humans do, enabling them to interact with people more naturally and diagnose mental health issues like depression.

Help Received
All project work was independent, but my mentor answered questions and gave advice. I also used various papers by Google related to CNNs to advise me in any architectural changes I made to my network ensemble in testing.

CALIFORNIA SCIENCE & ENGINEERING FAIR
2018 PROJECT SUMMARY

Name(s)
James L. Wang

Project Number
S1520

Project Title
Improving Mitosis Detection and Localization in Breast Cancer Histology Using Deep Convolutional Neural Networks

Abstract

Objectives/Goals
To (1) create an enhanced machine learning model for classification of mitosis in breast histology through transfer learning on pre-trained deep convolutional neural networks and (2) to develop a more effective and accurate algorithm for localizing instances of mitosis in breast histology instead of using recurrent convolutional neural networks.

Methods/Materials
Open-source histology slides were first obtained from the MITOS-ATYPIA 14 Grand Challenge These slides contained 391 instances of mitosis and 741 instances of false-positives, which were extracted through image segmentation techniques. Data augmentation techniques were then used to artificially augment the dataset to 3,000 images, with 1,500 in each class. These images were then fed through a convolutional neural network for training using MATLAB and an NVIDIA GTX 1060 graphical processing unit. Using an optimal trained model, a selection-search based algorithm was developed to more accurately localize mitosis instances in breast histology.

Results
After training this dataset on different transfer learning implementations on well-known pre-trained convolutional neural networks, it was found that retraining the final fully-connected layer and softmax classifier of the VGG16 model yields the best performance, with a 91.33% classification accuracy on a reserved test set. Our selection-search based algorithm was able to localize most instances of mitosis on histology images; however, some false positive instances were also detected by our algorithm.

Conclusions/Discussion
Our histology analysis pipeline is able to localize and accurately classify instances of mitosis in whole-slide images. While we did find limitations in our localization algorithm, they can be improved by allowing the reading kernel size to be adaptive to the relative sizes of mitotic instances in each whole slide image. Our algorithm design examines the problem of localization with emphasis on accuracy rather than efficiency, which has been explored less in this field of research. The methodology used in this research can be easily generalized and applied to other medical imaging tasks.

Summary Statement
A breast histology analysis pipeline was developed to more accurately detect instances of mitosis using transfer learning on convolutional neural networks and by developing a selection-search based localization algorithm.

Help Received
Breast histology slide images were acquired from the MITOS-ATYPIA-14 Grand Challenge Image Dataset. Neural network transfer learning implementations and algorithm design were done independently without external guidance.