CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Naomi K. Berhane

Project Number

S1401

Project Title

The Mysteries of Ancient Mathematics

Abstract

Objectives/Goals

The purpose of this project was to analyze an Ancient Ethiopian method of multiplying and relate it to binary arithmetic. This method skillfully avoids using fractions and the multiplication table which was ideal for the uneducated merchants that frequently used it with coffee beans. When multiplying two numbers, for example 25 and 31, the method divides 25 by two, while ignoring any remainders of one, until it reaches one. This is done in one column while in the second column, 31 is doubled as many times as 25 is divided by two. In any row that has an even number in the left column, the entire row is discarded. Then the entire column with the divisions of 25 is discarded. By adding the remaining groups of numbers, the correct product is given. My hypothesis is that the Ancient Ethiopian Method of multiplication is equivalent to the modern method used by computers based on the binary system.

Methods/Materials

This connection was made by first understanding how binary math works then relating it to steps of decimal multiplication and the Ancient Ethiopian Method of multiplication.

Results

The hypothesis was proved by discovering that dividing and doubling the numbers was converting a number to binary, and making the other number recognize that binary factor.

Conclusions/Discussion

The fact that this complex binary system, which is used in modern day computers, was used over 2,000 years ago shows how advanced some early civilizations were. I concluded that the method is quite general in the way that similar charts for decimal math and other bases such as base7 multiplication can be created

Summary Statement

Analyzing an ancient method of multiplication and explaining why it works.

Help Received

My father reviewed gramatical errors on my report.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)
Nikhil Cheerla

Project Number
S1402

Project Title
Accurate Detection of Skin Cancer Using Multi Stage Neural Networks

Abstract

Objectives/Goals
The objective of this project is to provide doctors an automated and accurate method to diagnose skin cancer from dermoscopic skin lesion images. This system must exceed 95% accuracy and must be able to learn from mistakes to improve its accuracy. The entire process must be packaged in an application that is simple to use.

Methods/Materials
This project has four major sections; image segmentation, feature extraction, neural network analysis, and developing desktop/mobile applications. I developed a two-pass image segmentation algorithm to segment the skin lesion from the surrounding skin. I developed techniques to extract a comprehensive suite of fourteen features to describe each lesion. I then trained a 2-stage artificial neural network classifier with these features extracted from a database of skin lesion images. I fundamentally changed the structure of how inputs were fed to the classifier, vastly increasing the performance (see results). Lastly, I created a desktop and a mobile application that allows doctors to use my research in practice. In this project, I used skin lesion images from a University of Chile database. I used MATLAB for image processing and neural network capabilities and a DermLite# DL1 dermatoscope for testing my mobile application.

Results
My segmentation algorithm was able to successfully detect light colored lesions and discard partial lesion images. Theoretical analysis predicted that a #system of experts# hierarchy or a #second opinion# neural network would both be effective ways to improve performance. This was confirmed by the results achieved: average performance of a 1-stage neural network was only 78.6% compared to 96.1% and 96.7% for my proposed 2-stage classifier system. The accuracy of the system stayed above 95% on repeated training and testing. The sensitivity (probability that a tumor would be detected) was on average 96.7%. Experimentation with test subjects confirmed that no prior experience with MATLAB is needed to use the desktop and mobile applications.

Conclusions/Discussion
I was able to meet all the project goals. My system achieved a higher accuracy than targeted, and was able to learn from new data. The automated applications were simple to use. My research would ideally be used by doctors equipped with an IPhone and dermatoscope to get an instant diagnosis. This technology has the potential to revolutionize skin cancer detection and save lives.

Summary Statement
This project is about an automated and self-learning system that accurately diagnoses skin cancer from skin lesion images.

Help Received
My mentor, Ms. Frazier, provided guidance on how to analyze the results and she also proof read my research paper. My dad helped me prepare the project display board.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Min Jean Cho

Project Number

S1403

Project Title

Applying Bayes' Theorem to DNA Sequence for Identification of Pathogenic Bacteria

Abstract

Objectives/Goals

To develop an easy, simple method for identifying microorganisms based on their DNA sequences, Bayes' theorem was applied to DNA sequence analysis. It was hypothesized that the conditional probability of a DNA sequence from an unknown bacterial species being a member of a particular species could be the posterior probability, which could be estimated from prior probability and likelihood function using Bayes' theorem.

Methods/Materials

To test the hypothesis, 16S rRNA gene sequences of foodborne pathogens (eight bacterial species) were downloaded from NIH's GenBank (45 sequences from each bacterial species, 360 sequences in total) to construct a database. Bayes' theorem was used to estimate the posterior probability of a bacterial specie "Si" given an unknown sequence "Q", P(Si|Q) = P(Q|Si) X P(Si) / P(Q). To determine the likelihood, P(Q|Si), the DNA sequence "Q" was divided into words (k-size DNA sequence fragments), and P(Q|Si) was measured from the average probability of observing the word j from species Si, P(wj|Si). The prior probability, P(Si), and P(Q) were calculated from the database sequences.

Results

The size of word (k) affected values of P(Q|Si) and P(Q). The optimum size of word (k) was determined to be 39 nucleotides. All test sequences showed the highest P(Si|Q) values for the species to which they belong, which indicated that the developed method correctly identified the test sequences (accuracy = 100%).

Conclusions/Discussion

The hypothesized algorithm was proven to work in the experiments carried out with DNA sequences of bacterial species. Dividing the unknown DNA sequence Q into small-size words (wj) was especially important to determine P(Q|Si) and P(Q). An unknown sequence should be classified into the species with the highest P(Si|Q) value (rank-based identification), which indicated the most probable species among the species included in the database.

Summary Statement

Bayes¡# theorem was applied to DNA sequence analysis in order to determine the conditional probability that a DNA sequence from an unknown species belongs to a particular species.

Help Received

My father assisted me with researching, and my mother read and edited my writing. I would also like to thank Ms. Julia Newman for her valuable advice and suggestions.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Dibya Jyoti Ghosh

Project Number

S1404

Project Title

Implementing Novel Kernel Discriminant Analysis to Optimize Breast Cancer Diagnosis

Abstract

Objectives/Goals

The objective is to determine if pre-processing mammographic data with KDA (Kernel Discriminant Analysis) increases the accuracy of a machine-learning breast cancer diagnosis algorithm by increasing detections and removing false positives. This will result in the simultaneous advancement of the medical and machine learning fields.

Methods/Materials

Mammographic images of 50 normal and 50 cancerous patients were obtained from the DDSM. These files were processed, filtered, segmented, and distilled into a set of 11 points each. These points were then processed with Quadratic (QDA) and Linear Discriminant Analysis (LDA) and stored separately. Three different machine learning algorithms of varying efficiencies were set up, and trained with subsections of the data, to analyze the effectivity of the Kernel Discriminant Analysis pre-processing technique over a range of data sizes and base training algorithms.

Results

Among the two methods tested, quadratic and linear, Quadratic Discriminant Analysis provides the greatest accuracy boost. Trends in results were uncannily similar between the three base machine learning algorithms, and followed the same general pattern. With small-scale training data sets (10-60 subjects) , QDA provides a 60% accuracy boost over the control group, and LDA a 30% increase. Under training data sets with size over 60, QDA provides comparable accuracy as the control, but LDA deteriorates and a 15% decrease in accuracy.

Conclusions/Discussion

The data collected shows that quadratic discriminant analysis is a viable pre-processing step for current laboratory breast cancer diagnostics. The algorithm has a marked increase in accuracy with smaller data sets, and adds low overhead, enough to justify implementation into diagnostic programs.

Summary Statement

I created a preprocessing algorithm for breast cancer detection programs that increases accuracy by mutating the mammographic data to better fit the algorithmic model.

Help Received

UCSF for providing mammographic data and suitable analysis methods. My dad for helping me debug my program at times

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)
Amber J. Graham

Project Number
S1405

Project Title
Noisy Music

Abstract

Objectives/Goals
My objective was to determine if certain kinds of music should be classified as music or noise according to physics through a mathematical analysis of the waveforms of samples of music.

Methods/Materials
I used a computer, a software called "Audacity," a total of 90 songs from 9 different genres of music, a recording of a car's engine as a control, and a recording of a chord on the piano as another control. I analyzed the entire waveform of each sample of music, noting in my lab book the times of the song that were "musical" and the times that were "noisy." If the song had a mostly musical waveform, the song was determined to be music, and if it was mostly noisy, the song was determined to be noise. I then gathered my data and looked to see how many songs were considered noise and music, and from which genres those songs were from.

Results
I could not make any definite conclusion from the results I had when I organized the music by genre because some genres only had two or three songs that were music or noise, and I could not say that absolutely every song from that genre is either music or noise. I then organized my data by year that the songs were composed in. The results were definitely more significant, and it was easily seen that older songs were more likely to be musical. There were still some outliers, though, and I wanted to figure out why that was occurring. After doing more research, I then organized my results by how the song was mastered (either for loudness or not for loudness), and the result was extremely significant. Out of the seventy samples of music that were mastered for loudness, absolutely zero of them were musical.

Conclusions/Discussion
Whether the song was mastered for loudness or not plays the main role in deciphering what types of music is considered "musical" and "noisy."

Summary Statement
Through a mathematical analysis of a total of ninety songs, it was determined in my project that there are certain types of music that should be considered "noise" and not music according to math and physics.

Help Received
None

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)
James Gui

Project Number
S1406

Project Title
Simulating Tumor Progression Using Spatial Evolutionary Games

Abstract

Objectives/Goals
My goal was to improve upon the model of tumor progression described in "Evolutionary game theory elucidates the role of glycolysis in glioma progression and invasion" by D. Basanta by incorporating a spatial component.

Methods/Materials
A. I first replicated results gathered in the original Basanta model by using an ordinary differential equation coded in Matlab. Then, by using the Laplacian matrix of a 10 by 10 grid, I attempted to turn the original model into a two-dimensional model.
B. I changed the parameters c, k, and n was well as the initial values and ran simulations to find out what changes in microenvironment could affect final cell populations.
C. The model was changed twice after creating replicating the original; once to add the spatial component, and once to improve visualization
D. I recorded the population fractions at the end of each simulation under different parameter values. For example, at c=.5, n=.4, and k=.2, the values of AG, INV, and GLY were 18.58%, 50.09%, and 31.33% respectively.

Results
The model returned similar results to the original model, but in a more clearly visualized manner. The microenvironments that were conducive to GLY cells were also conducive to INV cells, and GLY cell takeover always preceded INV takeover.

Conclusions/Discussion
The model was an improvement on the original model in that it provided a more solid visualization and accounted for space as a factor in tumor progression. However, the assumptions surrounding the simulation could be adjusted in accordance with more specific biological background.

Summary Statement
I use MatLab to improve upon an existing model of glioma progression by including space as a factor in analysis.

Help Received
Graduate student helped with MatLab syntax and initial background information

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Agastya Gupta

Project Number

S1407

Project Title

An Inexpensive, Global, and Effective Means of Diagnosing Heart Disease via Computer Imaging

Abstract

Objectives/Goals

As per the World Health Organization factsheet No. 317, cardiovascular diseases, are the number one cause of death around the world and represent 30 % of deaths worldwide. Furthermore, The Centers for Medicare & Medicaid Services (CMS) report that further hospitalization for deteriorating heart conditions causes $12 Billion in potentially preventable healthcare costs in the US. The Jugular Venous Pulse (JVP) is one of the most critical indicators for cardiovascular diseases.

Currently, measuring JVP accurately is a complex and expensive process fraught with errors, requiring highly experienced physicians to differentiate between the jugular venous pulse and the carotid arterial pulse, identify the sternal notch, and use unwieldy rulers to correctly take JVP pressure. I set out to develop computer vision based diagnosis software that can offer this same diagnosis based just on the video of the neck of a patient.

Methods/Materials

My software utilizes a customized MATLAB Computer Vision algorithm, combining Optical Flow and Blob Analysis to isolate the biphasic pulse rhythm of the JVP and determine its presence on the neck. I constructed and refined a customized Optical Flow and Blob Analysis algorithm, preceded by heavy morphological post-processing. I created a JVP simulator system and then used it to test the software in various lighting conditions to ensure its success under the actual lighting conditions of a user. Finally, the software was tested on actual hospital patients and correctly identified the pulse.

Results

The software was tested using the JVP simulator under environments with a shadow, without a shadow, and with speckle noise and could detect the pulse successfully more than 90% of the time over 30 trials. The software was then used on actual hospital patients and has successfully detected the pulse.

Conclusions/Discussion

The immediate results of the software will allow patients to monitor their heart health on a more regular basis and would allow for a rapid response to deteriorating symptoms. This solution provides an inexpensive, fast, and global means of monitoring and managing cardiovascular health in even the remotest parts of the world. My software is patent pending.

Summary Statement

I created a software that utilizes a customized Optical Flow and Blob Analysis computer vision algorithm to detect the presence of the Jugular Venous Pulse (JVP) to diagnose heart disease from just a video of the neck.

Help Received

Recorded patients at the Stanford University School of Medicine during a paid internship with Professor Paul Wang. Feedback on software design and results by Professor Paul Wang and Dr. Jeffrey Caves at Stanford University. My parents helped with logistics and printing of documents.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

William C. Hang

Project Number

S1408

Project Title

Semantic Multilayer SVM: Novel Artificial Intelligence Applied to Prostate Cancer Grading and Breast Cancer Diagnosis

Abstract

Objectives/Goals

We develop a novel Support Vector Machine (SVM) learning algorithm called Semantic Multilayer Support Vector Machine (SMLSVM) that can take groups of low-level numerical data and translate them into higher-level concepts. We also develop an image analysis toolkit, Procist, to translate a prostate histology image into numerical data. We apply Procist and SMLSVM to prostate cancer grading based on two Gleason Grades, a novel development.

Methods/Materials

SMLSVM generates fewer, more valuable higher-level features out of many, lower-level numerical features. These capabilities allow it to reduce the curse of dimensionality and the semantic gap.

Procist is a computer vision algorithm that can generate numerical data that characterize the spatial arrangement and irregularity of cells, nuclei, stroma, lumens, and other tissue elements.

All algorithms were developed in MATLAB, and were tested on 20 prostate cancer histology images from the Johns Hopkins Medical Institute (JHMI) without patient info. We further validated SMLSVM alone on the University of Wisconsin Breast Cancer Dataset to differentiate between 699 malignant and benign tumor cases.

Results

We performed 10-fold cross validation on both datasets, and preliminary testing indicates that my algorithms outperformed current SVM approaches on both datasets. Results from further testing will be updated and presented during the fair.

Conclusions/Discussion

The objectives of this project have been achieved. SMLSVM and Procist could classify cancer malignancy and cancer severities with very high accuracies.

Few studies have been performed on Multilayer SVMs, and we present new research into SVM technology, artificial intelligence, and deep learning. Furthermore, by differentiating cancers based on two Gleason Grades, we present novel contributions to computational histopathology, where current approaches only differentiate between one Gleason Grade and are much less clinically accurate.

A provisional patent is pending on SMLSVM. I am currently working with JHMI to secure a much larger dataset for a better validation of my algorithms' capabilities.

Summary Statement

We present a novel artificial intelligence architecture that uses the collective power of SVMs to perform data abstraction, and apply this method along with newly developed image processing algorithms to autonomously grade prostate cancers.

Help Received

Dr. Robert Veltri from Johns Hopkins Medical Institude supplied necessary testing images and support, David West provided encouragement and advice.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Kevin A. Hsieh; Abrahim N. Razzak

Project Number

S1409

Project Title

Optimizing Quantization Matrix Scale Factor and Pixel Density for Effective Optical Character Recognition in JPEG Images

Abstract

Objectives/Goals

The purpose of this experiment was to determine the effect that changes in quantization matrix scale factor (abbreviated "S-factor") and pixel density have on the effectiveness of optical character recognition (OCR) in JPEG images, and to create and use a script for finding optimal values for those variables under a given set of conditions. It was hypothesized that OCR effectiveness increases (1) as pixel density increases and (2) as S-factor decreases, because increasing pixel density improves clarity, but increasing S-factor introduces compression artifacts.

Methods/Materials

This experiment requires a computer capable of running batch files. Other software programs required included IrfanView and Tesseract OCR. In order to test the effect that the manipulated variables of S-factor and pixel density have on OCR, a script based on batch, a scripting language used to coordinate programs with large numbers of files, was created. The script was run three times, on three different sets of words, for three trials. This resulted in 29,700 data points, which were then tabulated and converted from correctly transcribed word count to a percentage representing OCR effectiveness. Visuals were created from the data for analysis purposes.

Results

In most trials, OCR recognized at least some words when pixel density was at least 40 ppcm, and recognized all of the words once the pixel density reached approximately 75 ppcm. A large amount of deviation was observed between 40 and 75 ppcm, but overall, OCR effectiveness improved as pixel density increased. The effect of Q-factor was unclear at first; data points had to be re-expressed in terms of S-factor for statistical analysis. It was observed that S-factor and OCR effectiveness were inversely proportional.

Conclusions/Discussion

The results supported both parts of the hypothesis and the associated justifications. This experiment assists in understanding what combinations of S-values/Q-values and pixel density produce useful OCR results, a concept that is relevant when determining optimal compression ratios for images intended to be processed with OCR, a process common in modern smartphones. Under the conditions of this experiment, a JPEG image at about 75 ppcm with an S-value of about 1.67 (Q = 30) would have allowed for maximum OCR effectiveness at a small file size.

Summary Statement

This project determined the effect that changes in S-factor and pixel density have on the effectiveness of OCR in JPEG images, and created and used a script for finding optimal values for those variables under a given set of conditions.

Help Received

Mr. Antrim advised and oversaw progress; Mr. Tipper helped with statistical analysis; Parents provided equipment and locations for testing

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)
Abraham N. Jellinek

Project Number
S1410

Project Title
Ex-HTML

Abstract

Objectives/Goals

Create an alternative syntax for HTML called Lima, modeled around the LISP family of programming languages. HTML is quite verbose for many tasks, and that#s what I wanted to improve, while allowing Lima to be compiled down to HTML. Similar projects have been created for CSS (Sass, LESS) and JavaScript (CoffeeScript), but HTML didn#t yet have anything similar.

I chose LISP because its syntax is versatile and simple, allowing the same syntax to be portable to HTML, CSS, and even JavaScript.

Methods/Materials

The Lima prototype was written in the Scala programming language, which includes a parser combinator library that was an integral part of the project. Additionally, it used a Java library called JTidy for cleaning up outputted HTML, the RSyntaxTextArea Java library for creating a live-updating GUI, and the ScalaJS compiler to create a web interface. The syntax was based on LISP S-expressions, but it used Clojure-style vectors for tag properties.

The final project is written in the Clojure programming language itself, without the use of any external libraries. The code is considerably better than the original, although it lacks JTidy integration, a GUI, and a web interface. I plan to add those features.

Results

The Scala version came out to around 400 lines of code. The parser was the slowest part, taking about 80 milliseconds to parse a 50-line document with complicated syntax. The HTML generator was faster, taking about 18 milliseconds to generate the HTML for that same document.

The Clojure version is much shorter and much faster: at around 130 lines of code, it parses and generates the same document in less than 10 milliseconds (usually closer to 5) and has many features that the original lacked.

Lima#s syntax is fairly simple, but it can be very powerful.

Conclusions/Discussion

Lima tends to be at least as readable as HTML, but requires much less typing. The compiler is fast, even for larger projects, and the syntax error messages are informative. It generates HTML5-compliant output, provided the user doesn#t make use of any deprecated tags.

Anything that can be written with HTML can also be written with Lima, and conveniences like macros and inline Clojure evaluation make creating web pages easier than before. I think that Lima is useful to web developers seeking refuge from the verbosity of HTML. JavaScript (through ClojureScript) and CSS

Summary Statement

An alternative syntax for HTML, focused on usability, extensibility, and consistency.

Help Received

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)
Tanisha Joshi

Project Number
S1411

Project Title
Quantitative Analysis of Macrocellular Biomarkers in Early Stage DCIS Cytopathology Images, Using Machine Learning

Abstract

Objectives/Goals

The higher incidence of DCIS cases in underdeveloped nations and an increased need for cytopathological tests have placed an undue learning burden on the medical enterprise to process millions of images in an efficient and cost-effective manner.

The purpose of this experiment is to write a self-learning algorithm with a preset confidence that is biologically aware of DCIS micropapillary/papillary disease in early stages.

Can we measure carcinogenic tumor activity in early stage DCIS using a machine learning technique with an 80% success rate?

Methods/Materials

I learned that cancer biology is intrinsically linked to histopathological outcomes. The extent of my work involved collecting images of papillary/micropapillary DCIS. The experiment analyzes 30 JPEG image files using ImageJ. (Training data set size = 500 image files)
The values for observed coverage are an average of 9 iterations per image.

Results

I was able to measure carcinomic activity in the ducts with an 80.9% level of accuracy from my algorithm. This measurement helps the oncologists focus on relevant DCIS images.

Conclusions/Discussion

My algorithm was able to learn how to identify ducts and measure the amount of carcinomic activity in them, by using an approach I invented and called Radial Vector Metrics (RVM) based machine learning.

The results met the engineering goal, because I was able to achieve 80.9% rate of accuracy. This is a valuable insight for future scientists because this research is proceeding in the right direction using an innovative approach (RVM). My results are specific to a particular sub-classification of DCIS-Micropapillary and Papillary carcinomas. I learned that I could consistently and reliably obtain results that showed the measure of carcinomic activity in low-grade DCIS.

Summary Statement

My algorithm measures carcinomic activity in the ducts with an 80.9% level of accuracy and this measurement helps the oncologists focus on relevant DCIS images and reduces their workload by minimizing human error.

Help Received

I would like to thank my teacher, Mrs. Mary-Kate Lesoine, for sponsoring my project. Dr. Hyma Vempaty, a hematologist/oncologist at Kaiser Permanente, validated my results. Mrs. Archana Gangakhedkar, an industry drug reseracher, also validated my results.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)
Sarah H. Kazmie

Project Number
S1412

Project Title
Smart Code

Abstract

Objectives/Goals

Modern credit card terminals store and transmit a purchaser's account information using standard computer protocols, which can be hacked and intercepted. In several recent attacks, hackers have used simple key loggers and packet sniffers to intercept millions of transactions.

My goal was to develop SMART CODE:
A Secure, Merchant-coded, Augmented Recording and Transmission using Character Optimized, Dynamic Encoding.

Methods/Materials

I developed a variation on the Huffman compression routine, where a sorted frequency table is weighted and skewed to produce a unique variable bit-length encryption. I created a function to artificially weight the symbol frequency table based upon the digits of one or more simple pin codes.

To test my algorithm, I created a C# application which can automatically compare and quantify the results of my compression algorithm on varying sets of multiple multi-digit code variations.

I also converted my SMART CODE algorithm into C and programmed it into a PSOC 4 dev board to demonstrate that it could be embedded in a micro-controller.

Results

I tested more than 990,000 samples of two three-digit pin codes, and more than 6 million samples from a random progression through three four-digit pin codes. In the more than 7 million total test comparisons, no messages could be fully decoded with even a single digit change in any of the pin codes. I wrote a function to quantify the difference in two resulting strings and not one of the 7 million sample comparisons, scored a difference value of less than 20,000, where 0 indicates identical text and 50,000 is completely random.

Conclusions/Discussion

SMART CODE successfully executed with three four-digit pin codes in less than 0.07 seconds. A continuous run of 990,000 samples, on a fast PC, took about 12 hours to complete. This represents the time it would take to find a match using all combinations of only 2 3-digit codes. If you had to try all combinations of two four-digit codes, it would take 100 times longer, and if the series were expanded to to include a third four-digit code, it would take about 12,000,000 hours or 1,369.8 years.

Summary Statement

An adjusted weighted Huffman Coding algorithm can be used to create a very secure and efficient encrypted credit card processing system.

Help Received

A family friend loaned me a PSOC dev board and taught me how to use it.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Janel (Jihyeon) Lee

Project Number

S1413

Project Title

An Enhanced Method for Fusing Multiple Exposure Images for High Dynamic Range Scenes

Abstract

Objectives/Goals

The goal of this project was the design and construction of a technique that effectively removes blurring and ghosting artifacts in the process of high-dynamic-range (HDR) imaging. The project was targeted toward handheld devices with cameras, including smartphones, which do not have the same capacity as large computer processing units (CPUs) and graphics processing units (GPUs). This improved technique takes HDR images that compensate for changes due to brightness or movement in a scene, producing images of a better quality than currently available.

Methods/Materials

Using a computer, a program to run the enhanced algorithm was created using Open CV and Matlab. Data sets were found and chosen to compare existing methods to the enhanced method. A camera was also used to test the algorithm on an actual device.

Results

A better algorithm was developed to successfully remove artifacts that arise due to scene motion and brightness changes in a scene. Without increasing computational complexity, it outperforms other methods by using contrast differences calculated with a Laplacian operator and Poisson blending to adjust images before they are fused.

Conclusions/Discussion

This project uses the power of HDR imaging to address human error in the use of technology. For point-and-shoot cameras, camera shake and brightness change are common problems, which in turn creates artifacts in photos. This project presents an enhanced solution to remove blurring and ghosting artifacts, producing better results without having to use more computational power. The improved method not only resolves the problems but provides a simplified, more effective solution to be used in handheld devices.

Summary Statement

This project resolves human error in HDR imaging by removing artifacts that are caused by scene motion and brightness change without increasing the complexity of the process, making it fit to be implemented in handheld devices.

Help Received

Mother helped glue things on board

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)
David K. Legg

Project Number
S1414

Project Title
Computerized Deductive Reasoning Using Bipartite Rules

Abstract

Objectives/Goals
My objective is to determine if automatically manipulating bipartite rules is an effective way to perform deductive logic.

Methods/Materials
A self-modifying system of rules was set up, with support for #grid puzzles# added. 36 puzzles, varying in size and difficulty, were selected at random and translated into an appropriate form for the program to interpret. The computer ran the algorithm for each puzzle, and data about completeness, accuracy, and time taken to reach a conclusion were recorded.

Results
The algorithm never made an incorrect deduction, though it sometimes failed to make enough deductions to complete a puzzle. It solved 61% of puzzles completely, averaging 4.5 seconds per puzzle.

Conclusions/Discussion
The algorithm successfully applied deductive logic, reaching correct conclusions in times that were reasonable. The bipartite rule structure is an effective way to implement deductive logic. Furthermore, it mirrors the way humans perform comparable logic tasks, as they use similar structures naturally. I plan on adding functionality to handle transitive logic and other more complex manipulations, which should increase the solve rate, especially for more difficult puzzles.

Summary Statement
I created an algorithm that manipulates bipartite rules according to the conventions of formal logic in order to perform deductive reasoning.

Help Received
Discussed ideas and issues with father.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)
Christopher Lu

Project Number
S1415

Project Title
Complete 3D Projection Using Head-Tracking and Stereoscopy

Abstract

Objectives/Goals
The objective is to construct a virtual image display system that incorporates head-tracking, stereoscopy, and other depth cues, as well as 360 degrees of viewing to create a superior 3D viewing experience.

Methods/Materials
A laptop, webcam, Arduino microprocessor, motor shield, stepper motor, and bearing were used in this project. The viewer#s head location was estimated by a face-tracking program based on the webcam#s video. With my C# code, the data was filtered and a 3D model was projected onto the screen via a projection matrix to simulate motion parallax and head-tracking. My programs, written in C# and Arduino programming language, also control and coordinate the image shown on the monitor using Unity, as well as the rotation of the monitor using a stepper motor so that it faces the user. Stereoscopy, the effect used in movie theaters, was added by using anaglyphs filter to simulate parallax as well.

Results
The multiple parts of the project, including the face-tracking program, the Arduino, and the anaglyphic view, in concert such that the effect is a completely virtual, hologram-like projection from the screen. Overall, the project is superior to many current technologies in that it is cheaper, incorporates more depth cues, has a larger viewing range, and allows for more interaction.

Conclusions/Discussion
3D technology is the next step in virtual display, giving it a lot of future potential in many fields. Although the most obvious applications are in entertainment and computer graphics, it is also usable in medical diagnostics, flight simulations, education, biomechanical studies, scientific visualization, weather diagnostics, chemistry, and many others.

Summary Statement
Several 3D depth cues are mimicked using a variety of techniques in order to create a hologram-like projection that is both innovative and superior to current technology.

Help Received
Father helped arrange the poster and buy the materials

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)
Nitya Mani

Project Number
S1416

Project Title
Characterizing the n-Division Points of Genus-0 Curves through Straight Edge and Compass Constructions

Abstract

Objectives/Goals
This project explores the field of constructibility of n-division points (points dividing a closed curve into n pieces of equal arc length) through straightedge and compass constructions, by furthering the work done by the two major theorems in this field - the Gauss-Wantzel Theorem and Abel's Theorem on the Lemniscate. Three major problems in this field were investigated: determining a closed form solution for the regular polygons that can be constructed with a straightedge, compass, and trisector; finding the values of n such that the n-division points of the tricuspoid can be constructed with a straightedge and compass; and seeking a generalization of Abel's Theorem to the entire family of Serret curves (a family of curves including the lemniscate whose arc lengths share properties).

Methods/Materials
Two major mathematical fields were used to algebraically represent the problem of constructibility. Field theory was used to characterize the figures, angles, and lengths that are constructible under a given set of conditions, and the theory of elliptic integrals was used to determine expressions for the n-division points of these curves. Three areas of mathematics were used in the proofs of the theorems. Galois theory was used to find a solution to the first problem, algebraic geometry was used to relate geometric problems of constructibility and n-division points to algebraic ones, and complex analysis was used to examine the elliptic integrals that characterized the arc length of these curves, particularly when solving the third problem.

Results
Three major results were obtained through the research. A closed-form solution for the values of n for which the n-division points of a circle can be constructed with a compass, straightedge, and trisector was found; a theorem was proved that for all integer n, the n-division points of the tricuspoid curve are constructible; and it was determined that with a compass and a straightedge, arbitrary arc lengths on any Serret Curve can be added, subtracted, and multiplied.

Conclusions/Discussion
These results represent the product of a year of investigation, however, additional work is being done to explore related problems in this field, such as examining the n-division points of a circle constructible with a straightedge, compass, and p-sector, as well as characterizing the n-division points of other significant closed curves.

Summary Statement
This research project explores the field of n-division point constructibility with straightedge and compass constructions through three problems, characterizing the n-division points of the circle, tricuspoid, and Serret Curves.

Help Received
Worked with Dr. Simon Rubinstein Salzedo from Stanford who taught me much of the background for my research; Mr. Spenner, my science teacher, sponsored my project to Synopsys.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Kai E. Marshland

Project Number

S1417

Project Title

Dart vs. JavaScript: A Benchmarked Comparison of Two Client-Side Languages

Abstract

Objectives/Goals

The project compares Dart and JavaScript, two programming languages, using the Kraken Benchmark.

Methods/Materials

Dart and JavaScript are both client-side languages; they run on the user's computer when a website is loaded. They allow for interactivity within websites. Dart is Google's replacement for JavaScript, as JavaScript contains numerous problems. The Kraken Benchmark is a series of tests, approximately 100,000 lines of code long, which is typically used to measure the speed of JavaScript in various browsers. The investigator translated the Kraken Benchmark into Dart, then ran the benchmark in both Dart and JavaScript 25 times. The speeds were then compared.

Results

While it was hypothesized that Dart would be much faster than JavaScript, Dart was in fact much slower, taking almost four times as long to run the tests.

Conclusions/Discussion

Dart is not yet ready for widespread use. This has real-world implications, as everybody is looking for ways to maximize browser speeds, especially on mobile devices.

Summary Statement

The project, by comparing the speeds of Dart and JavaScript using the Kraken Benchmark, which the experimenter translated into Dart, demonstrated that Dart is not yet ready for widespread use.

Help Received

Father helped edit writeup

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Jeffrey Ni; Mitchell Wu

Project Number

S1418

Project Title

Automating Emotion Recognition for Music

Abstract

Objectives/Goals

Finding the right songs to use in a video is a tedious process, as one must manually listen and evaluate many possible songs to determine whether they express the right moods that fit the video. An easy-to-use tool for video producers to analyze the emotional content of songs, search for them, and visualize moods within them is presented in this project.

Methods/Materials

We assessed two primary approaches for determining the mood of the song: lyrics analysis and audio analysis. For lyrics analysis, a custom algorithm that looks for specific keywords from the ANEW dataset was implemented in Java and tested. For audio analysis, moods were determined by a classifier created with the WEKA machine learning toolkit, along with a ground truth dataset for the classifier to be built upon. Features of the audio were extracted using jAudio and correlated with the manually classified mood we gave the song to form the ground truth. Several classifiers that WEKA supports, such as Support Vector Machine and Naïve Bayes, were tested before we selected a specific one to use. The results of the song mood analysis are then stored in an Oracle database, which can be searched with a graphical tool.

Results

Between lyrics and audio analysis, we found that our audio analyzer using Naive Bayes was the most effective. The accuracy of whole song audio analysis, as determined by 10-fold cross validation, was around 80%. Compared to manually analyzing and finding songs, our software automatically classifies songs by the moods that they express, helps video producers find songs quickly, and allows users to immediately see the moods throughout a song.

Conclusions/Discussion

The primary engineering goal has been met. Further work could be done to improve the accuracy of the lyrics analysis. With lyrics and audio analysis combined, an even more accurate classifier can be made.

Summary Statement

We created a prototype music analyzer that enables video producers to classify music according to emotional content, search for music, and visualize the emotions in order to determine if it suits the emotional context of a video.

Help Received

Mitchell#s dad taught us about relational databases and some Java programming. He offered suggestions when we had difficulties, but we ultimately made the final call on any decision.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Jongho Park

Project Number

S1419

Project Title

The Automatic Detection of Comets through Image Processing LASCO C2 of the Solar and Heliospheric Observatory

Abstract

Objectives/Goals

Since the Solar and Heliospheric Observatory (SOHO) started its mission in 1996, using the coronagraphs from the Large Angle and Spectrometric Coronagraph (LASCO) C2 and C3, many amateur astronomers found comets by comparing the movement of potential comets to the movement of other celestial objects by eye. This project is aimed to develop a program and algorithm to automatically detect comets, which appear in the image of LASCO C2 of SOHO.

Methods/Materials

First, I would brainstorm and code a program in MATLAB to extract celestial objects in the image set and then distinguish comets from other objects, such as stars and cosmic rays. Next, I found data of known comets and input the images of the known comet in order to test the prototype. Also, to determine certain constraint variables, the data of known comets were gathered and statistically analyzed The function of the program was redesigned through major and minor changes during coding. When all errors were fixed, to test accuracy, 25 stratified random samples of comet containing images were selected through the years 2011-2013 with known comet orbits (x and y coordinates).

Results

While minimizing the threshold value yielded the comet, when it exists in the images, it greatly lowered the accuracy and efficiency due to false results and long calculation time. In order to increase accuracy, the comet-detecting program, or function, needed a greater number of input images, while lowering the threshold value. This approach maximized accuracy, when there were seven image parameters and threshold value equaled 0.06, to 46.67%, which contains the information of a comet among output orbit data by 46.67%.

Conclusions/Discussion

The program developed could yield a maximum accuracy of 46.67%, which means among 100 of the coordinate outputs, 47 of them would be comets. Although this program may not extract comets only, the algorithm significantly reduces the observations that would take place without the program and produce a list of potential comets. For further research, the accuracy can be enhanced by adding the condition of comparing RGB values for similar brightness.

Summary Statement

This project is aimed to develop a program to automatically detect comets, and create a larger database of comets, leading to more sophisticated research of comets, and algorithm that may be applied to other fields of astronomy.

Help Received

I would like to thank Dr. James Li and Mr. Knight for their help for me to start and enter science fair and discussions.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Kalyani Ramadurgam

Project Number

S1420

Project Title

High Dimensional Clustering Algorithms Applied to Face Recognition of Obscured Faces

Abstract

Objectives/Goals

The objective of this project is to develop a method for identifying faces in situations where a face is angled away from the camera or partially obscured. Novel recognition of hidden faces is urgently needed in settings of both military security and personal use. Using high dimensional clustering algorithms like KNN, this project aims to increase the accuracy and versatility over current facial recognition software as well as run on devices that can be used in daily life.

Methods/Materials

I used parts of the CMU Face Images Dataset as training and testing data and OpenCV libraries to implement parts of the preprocessing steps. I processed and filtered all pictures in the dataset with Gaussian blurring, Dodge high-contrast filtering, and greyscale conversion. To crop the faces and extract facial features and ratios, I used the Viola Jones Algorithm rejection cascade. The KNN clustering algorithm was used as the main machine learning algorithm to calculate the pictures closest together in a high-dimensional hyperspace.

Results

Using the complete CMU dataset with both obscured and unobscured faces, the system generated an accuracy rate of 90.4%, which is very comparable to modern use. When faced with only obscured faces, the algorithm had an accuracy of 55%. Current techniques have proven to be almost completely random when only given hidden faces, which is about a 4.95% accuracy rate. So, the algorithm presented in this project provides almost a 50% increase in accuracy over current methods when working with hidden faces. Further analysis shows that pixel mapping carried the weight of the system when compared to facial features.

Conclusions/Discussion

I have successfully created a face recognition system that recognizes obscured faces with a higher accuracy than current techniques. The combination of pixel mapping and feature extraction makes it possible to classify faces that are facing away from the camera, wearing sunglasses, or are hidden from view. This is applicable to areas such as biometric security, crime identification, military security, and many other fields. As a result, this project has the potential to make the world a safer and more functional place.

Summary Statement

By combining pixel mapping and feature extraction, I created a successful and powerful facial recognition system that accurately recognizes angled and obscured faces, and is applicable in both military security and personal identity.

Help Received

Thanks to Samvit Ramadurgam for inspiring me and guiding me in my search for the best algorithms and approaches.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)
Sonia Sachar

Project Number
S1421

Project Title
Adaptive Visual Spatial Representation Based Trigonometric Learning System

Abstract

Objectives/Goals
This project addresses the need to enhance math learning to prepare students in secondary education in STEM. The project was created to teach trigonometry by incorporating visual spatial diagrams using the unit circle with a modified spaced learning algorithm to address each student's individual learning needs.

Methods/Materials
The application was created on a mobile platform as it is used by high school students in this sample group for the study. A lesson plan was devised to implement a self-learning algorithm to teach a subset of identities among trigonometric expressions involving the sine and cosine functions. To compare different types of learning methods, four learning systems were created: rule-based, rule space-based, visual spatial based and space-based visual spatial. A lesson test was designed and implemented to run pre-tests and post tests, and performance results were tabulated.

Results
The results proved that the post-test scores were thirty-seven percent higher in users who used the visual spatial space-based learning method compared with control group users who used the rule based method.

Conclusions/Discussion
In conclusion,these results from the experiments are significant in understanding how different learners can be taught and how adapting a specific learning style to each individual improves one's performance. Moreover, when visual spatial learning is combined with space-based learning the best retention of these trigonometric concepts is achieved.This system of learning through a utilized mobile application platform is economically feasible and technologically advanced.Implementation of this learning system in classrooms as an enhanced learning tool in trigonometry will address the learning needs of all students, especially visual spatial learners. Thus, students will be better prepared for STEM education careers.

Summary Statement
My project teaches high school students pre-calculus trigonometric concepts on a mobile application using a visual spatial learning approach.

Help Received
iOS Developer Program, Past Research Studies

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Anurag Singh; Kaushik Tandon

Project Number

S1422

Project Title

Parking Pigeon: Application for Enhanced Localization in Multi-Story Parking Lots

Abstract

Objectives/Goals

People forget where they parked their car in multi-story parking lots, and parking levels sometimes look alike. GPS systems do not work well inside buildings, or provide parking level information. Parking Pigeon, our Android application, uses built-in pressure sensors in smartphones to lead the user back to the correct floor.

GOAL: Design a smartphone application to locate the parking level of a car in a multi-story parking lot.

DESIGN CRITERIA

1. Easy to use smartphone app - guides user to correct parking level 2. Resolution better than 3 meters height of parking levels 3.Works in above and underground lots 4. No special equipment needed for the car

Methods/Materials

Parking Pigeon works in two steps. When parking the car, the app uses built-in pressure sensors to record the atmospheric pressure at the parking level. When locating the car, the app uses the pressure measurement at the current location of the user to guide them back to the correct parking level, by minimizing the pressure difference. To estimate the naturally occurring drift in atmospheric pressure from the time the car is parked to the present time, we query the regional weather station over the internet. This drift correction is applied to the current pressure measurement before converting it to altitude.

We used three smartphones for testing. Eclipse IDE was used for the Android development. We tested our application at 4 different multi-story parking lots on multiple days.

Results

Excellent correlation ($R^2>0.99$) between the height predicted by the pressure sensor and the actual heights of the parking levels. The average error of ~1 meter does not affect our results as typical parking levels are separated by 3-5 meters in height. In contrast, GPS data showed no correlation with actual height, and was inadequate for our purposes.

Our novel method of querying weather services for atmospheric pressure is successful in correcting for natural pressure drift. We have shown an excellent one-to-one correlation ($R^2>0.99$) between the regional atmospheric pressure with measurements from 3 separate smartphones on 4 different days.

Parking Pigeon was verified to work inside buildings and in underground parking lots

Conclusions/Discussion

Parking Pigeon is a successful prototype that met our engineering goal. It demonstrates an innovative problem solving algorithm using pressure sensors, weather stations and web query APIs.

Summary Statement

Parking Pigeon is a smartphone application to locate the parking level of a car in a multi-story parking lot, utilizing built-in pressure sensors.

Help Received

Mr. Prateek Tandon mentored on Android Programming. Mr. Pankaj Tandon and Mr. Ajay Singh helped with transportation and printing of the poster.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Yousuf M. Soliman

Project Number

S1423

Project Title

Indium: Using Novel Machine Learning Algorithms to Develop a Nondisease-specific Personalized Medicine Engine

Abstract

Objectives/Goals

In medicine today, treatments for patients are based on the diagnosis, not the patient individually. It has been shown that this is only effective for about 60% of people. To help treat more patients I developed Indium, a nondisease-specific personalized medicine using novel machine learning algorithms. I divide this problem into 3 distinct steps: diagnosis, prognosis and individualized treatment creation.

Methods/Materials

To truly create a completely generalized diagnostic system I developed a powerful natural language processing (NLP) engine that analyzes how language acts on itself. I then connect this NLP to PubMed, a database of medical research papers, to extract features and biomarkers that are indicative of certain diseases. Analyzing the effectiveness of my system, I have found that my generalized algorithm is more effective, by about 3 standard deviations, than the state of the art techniques and physicians. My prognostic software works by utilizing fuzzy lagged data co-clustering, an NP-complete problem. To circumvent this problem I developed a Monte-Carlo approximation that runs in polynomial time. Lastly, dealing with treatment creation I developed a Q-learning algorithm that dynamically adjusts for the specific patient parameters. To deal with the problem of censored data, I created an SVM system to maintain a constant belief state of the subject. I demonstrated the performance of the proposed algorithmic framework through the analysis of real clinical trials.

Results

The personalized medicine system I developed not only operates in non-optimal environments, but it is more effective than the state of the art techniques. Since my algorithm is offloaded into the cloud I am able to help patients around the world regardless of their socioeconomic status.

Conclusions/Discussion

If one can predict the likely result of a sequence of actions or treatments for some time out into the future, then they can use that to determine the optimal action right now.

The work presented here adds to a growing body of evidence that such complex treatment decisions may be better handled through modeling than intuition alone. This is true due to the fact that it has been shown to more accurately predict the optimal treatment plan than trained physicians.

Summary Statement

I developed powerful artificial intelligence algorithms to determine the optimal treatment path for specific patient by analyzing the entire patient history at an individual level.

Help Received

I consulted with several professors and physicians about the eminent need for personalized medicine.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Vivek Sriram

Project Number

S1424

Project Title

Using Agent-Based Modeling to Simulate the Transfer of Zoonotic Pathogens

Abstract

Objectives/Goals

Replicable experiments of zoonotic diseases are impossible due to the transient, chaotic nature of natural environments and the large scale of dynamics involved. I hypothesized that an agent-based model of the spread of malaria that simulates sufficient variable complexity and the new containment strategy of introducing genetically-modified mosquitoes would help effectively conduct experiments on possible treatment plans and accurately predict a lower human fatality rate than that for current treatments.

Methods/Materials

I designed a computer simulation backed by mathematical equations that models the spread of malaria from vectors to targets under various conditions. My model took several variables into account including temperature, humidity, amount of standing water, extent of urbanization, sickle-cell allele frequency and efficacy of current malaria control. Using a range of constants for each factor, I devised mathematical equations that calculated values for variables such as larval maturation rate, mosquito biting rate, and infection probability per bite, which were then used to determine population change rates for each group of agents: individuals that represent a certain class of mosquito or human. I also devised a possible treatment plan, where genetically-modified mosquitoes that cannot carry malaria were introduced into the population. Ultimately, I was able to run my model with 180 different combinations of variables. The graphs of infected humans over time produced were compared with actual infection rates, and the mathematical equations were refined to reflect realistic behavior.

Results

For a typical Kenyan city, my model predicted that 17% of the human population would be infected after a year, which is quite close to the actual proportion of around 20%. Analyzing the outputs I gathered for various combinations, I determined that my simulation is fairly accurate, with an average percent error of only 15%. With the addition of my treatment plan, my model predicted for the same Kenyan city that only 5% of individuals would be infected each year.

Conclusions/Discussion

The accuracy and efficiency of my agent-based model at predicting infection rates suggest that agent-based models offer tremendous promise as the tool of choice for determining the best eradication strategy for diseases transferred through intermediate vectors that involve complex, dynamic interactions.

Summary Statement

To evaluate the efficacy of simulations to study zoonotic diseases and eradication techniques, I created a complex agent-based model that accurately predicted the transfer of malaria and tested inclusion of genetically-modified mosquitoes.

Help Received

Mr. Sutton, Environmental Science teacher at The Harker School, recommended relevant reading and provided feedback.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Prem M. Talwai

Project Number

S1425

Project Title

An Investigation of the p53 Ubiquitin-Proteasome System Using a Novel Non-Steady-State Enzyme Kinetic Model

Abstract

Objectives/Goals

Existing steady-state models have failed to accurately describe the cancerous mechanism known as overzealous p53 ubiquitination because they rely on the quasi-steady-state assumption, which is invalid in the p53 ubiquitin-proteasome system (UPS). My project aims to develop a novel non-steady-state mathematical model of reversible sequential bi-substrate enzyme kinetics, which can be used to analyze the nonlinear dynamics of the p53 ubiquitination reaction for various initial concentrations of targeted p53, MDM2, and E2D3-Ub.

Methods/Materials

By exploiting the recurrence of certain rate terms in the conventional nine-dimensional mass action model of reversible sequential bi-substrate enzyme kinetics, a novel set of mathematical expressions were derived for the concentrations of each of the four intermediate complexes, which enabled the elimination of four superfluous variables from the existing model without the use of inaccurate steady-state or rapid equilibrium assumptions. The resulting five-dimensional mass action system was then linearized using conventional perturbation methods and subsequently solved using standard linear algebraic techniques. The model was then simulated in Mathematica to analyze the effects of E2D3-Ub and MDM2 concentrations on the rates of p53 ubiquitination at different temperatures.

Results

At each temperature, it was observed that the ubiquitin-ligase MDM2 accelerates the carcinogenic ubiquitination process, while ubiquitin-conjugated E2D3 inhibits it. It was also discovered that E2D3-Ub is a more effective inhibitor of overzealous p53 ubiquitination when present at higher concentrations. However, it was observed that high concentrations of p53 hinder the ability of E2D3-Ub to decelerate the reaction. The mathematical model successfully reproduced the experimentally observed p53-MDM2 interaction.

Conclusions/Discussion

The derived model therefore suggests MDM2 as a prospective target for cancer therapy. In addition, the findings of this project propose recombinant E2D3-Ub as a new promising protein-based anticancer drug for targeting overzealous p53 ubiquitination. The derived model can suggest new therapeutic strategies for targeting various neurodegenerative diseases characterized by an overzealous UPS. Finally, computational simulation of this novel model provides a safe, fast, and cost-effective preliminary alternative to expensive in vitro experimentation.

Summary Statement

My project derives a novel non-steady-state mathematical model of reversible sequential bi-substrate enzyme kinetics, which suggests E2D3-Ub as a new promising protein-based anticancer drug for targeting overzealous p53 ubiquitination.

Help Received

No help was received.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Arjun M. Tambe

Project Number

S1426

Project Title

Improving Algorithms for the Optimal Allocation of Security Resources, Year 2

Abstract

Objectives/Goals

Security is a ubiquitous concern, vital for counterterrorism, wildlife reserve protection, crime prevention, and many other critical applications. Strategic resource allocation is crucial since limits on security resources prevent full protection of all targets at all times. Since adversaries can exploit predictable security strategies, protection schedules must be efficient and random. Algorithms based on game theory offer a mathematically sound approach for creating weighted random strategies that account for predicted adversary reactions and different values of different targets, and they are currently being used to allocate security resources at several global locations. While many of these algorithms assume perfect rationality among adversaries, a recent algorithm, MATCH, accounts for humans' imperfect decision-making and has offered better protection than other algorithms. Last year's research developed a new algorithm, NewMATCH, that outperforms MATCH in certain cases. Both MATCH and NewMATCH contain a certain parameter whose value affects the operation of the algorithm, but changing this value has not yet been explored in detail. This project aims to create new procedures for adjusting the value of this parameter to make MATCH and NewMATCH more effective.

Methods/Materials

This study offers 2 innovations: a new model for predicting adversary behavior that is also used in a procedure that determines the optimal value of the parameters in MATCH or NewMATCH, and the application of this procedure to tune these parameters. To test the procedures as applied to both algorithms in a setting that models a real-world security situation, this study solicited online participation with human subjects. Subjects took the role of an attacker in an online game, playing against a security force whose strategy was determined either with or without the new procedures.

Results

The new model for predicting adversary behavior better predicted adversary behavior than any of the other existing models tested. Algorithms whose parameters are tuned via the new procedures are shown to be more effective than algorithms that are not tuned.

Conclusions/Discussion

The new procedures offer significantly increased security and improved predictions of adversary behavior. The innovations in this project may have great potential to reduce the risk of dangerous security breaches if applied in the real world.

Summary Statement

A program for tuning algorithms that allocate limited security resources and a new model of human behavior are developed, which are more effective than procedures currently being used to protect many vulnerable locations.

Help Received

Dad (computer science professor) advised the research and writing of the abstract.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

Tanay Tandon

Project Number

S1427

Project Title

A Machine Learning Model for Automated Semantic Short Essay Assessment through Random Forest Based Ensembles and NLP

Abstract

Objectives/Goals

This research focused on the development and training of a Machine Learning model for extracting meaning from natural language data. In our world today, it has become of increasing importance for computer models to intelligently understand and analyze textual data in natural language format. The model was taught to classify semantics from text through the use of Treebanks (Parse-Trees), Bag of Words Models, Sentiment Classification, and Structural Subject-Predicate patterns. The research was applied to teach an artificial-intelligence the task of autograding short essay responses by learning from teacher grading patterns.

Methods/Materials

The research was conducted primarily in Python through the use of the Natural Language Toolkit, sckit-learn RF package, and public training data from the Hewlett Foundation. Ten different essay sets with 17,044 responses ranging from the fields of chemistry to language arts were used for training the algorithm. The research hypothesized that through the use of semantic feature modeling and randomized ensemble learners, the learnt model could achieve a level of accuracy comparable to human grading accuracy # 0.80-0.85 Quadratic Weighted Kappa Inter-Rater agreement. The kappa score represents a level of agreement between a gold standard and a test standard for content evaluation. Development occurred over a period of 5 months, with focus on 4 distinct stages: Preprocessing, Feature Generation and Training, K-Fold Cross Validation, Final Model Packaging.

Results

In the Model Evaluation and Testing phase the algorithm received a Quadratic Weighted Kappa (QWP) average score of 0.76, which represents the inter-rater agreement between the model and the grades given by the human grader. This was based off of a separate test set of 6,000 elements and averaged over 10 cross validation sets.

Conclusions/Discussion

Overall, the developed model can be of valuable use for education practitioners and standardized testing in order to automatically grade large swaths of short written responses. The algorithm presents a near human accuracy in the autograding task (0.76 QWK). Furthermore, the research represents a level of text-intelligence and the model can be applied in a variety of military applications such as opinion mining, threat identification, and preemptive terror analysis.

Summary Statement

This research focused on training an artificial intelligence model to extract meaning from text and learn to accomplish the task of autograding short essay-responses through Random Forest ensembles.

Help Received

Parents and teacher provided advice on paper, Richard Socher from Stanford AI Lab provided answers regarding questions over email.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)
Marisa L. Thompson

Project Number
S1428

Project Title
Photo Identification of Chinook Salmon

Abstract

Objectives/Goals
The purpose is to use spot pattern recognition procedures and image processing to identify individual Chinook salmon using a program called Stripespotter. We must first test the accuracy of the program by comparing photos of different Chinook salmon by categorizing them in a specific ranking order.

Methods/Materials
Two buckets, anesthesia, camera, Styrofoam board, Photo Elements, Stripespotter
I removed three salmon from 15 tanks and placed them in a bucket, shortly before adding a few drops of anesthesia. I placed a label next to each fish that indicated which fish it was and the tank number it was in. After completing the 15 tanks, I repeated the process, while labeling each tank number with a letter B. With the 90 different photos of Chinook salmon, I extracted a portion of the spots from each fish using Photo Elements, and then entered the revised pictures into Stripespotter, which ranked these photos according to the "likeness" of the main photo that I selected.

Results
Ninety of the 90 photos listed as rank 1, while 22 of the 90 photos, or 24.4%, listed as rank 2, indicating a "complete accuracy." After examining further results, there was a steady decrease in the number of identifiable pictures as the rank number continued to increase until it reached rank 10. According to my data, 50 of the 90 photos, or 55.6%, were ranked within the first ten rankings, indicating a "partial accuracy."

Conclusions/Discussion
To some degree, Stripespotter is an accurate program; however, workers at hacheries shouldn't solely rely on a program that doesn't rank every original photo as rank 1 and every secondary photo as rank 2; again, it depends on that perception of "accuracy." Possible errors may have occurred in this eperiment such as fish movement or light quality, so before characterizing Stripespotter as an inaccurate program, we must first improve image processing.

Summary Statement
The purpose is the test the accuracy of a coding program called Stripespotter to determine if identifying Chinook salmon through photo recognition is a more effective way of tracking fish.

Help Received
Used equipment from the Merced River Fish Facility under supervisor Mr. Adelizi, received physical help e.g. lifting buckets, photographer took pictures

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)
Jared A. Tramontano

Project Number
S1429

Project Title
On the Theory of Lures with Dynamical Action on Compact Topological Manifolds and Ordinary Hyperreal Fractal Strings

Abstract

Objectives/Goals
In this project, we attempt to provide a new notion of a topological homeomorphism by defining such in terms of a dynamical system. That is, we want to construct a time-dependent quasi-attractor and a time-dependent basin of attraction that respects the homotopy equivalence class of some compact topological manifold by acting on some portion of the manifolds boundary, whilst altering the Minkowski Content.

Methods/Materials
There is no distinct set of steps one must follow in order to do research in mathematics. The "lure" was constructed through trial and error; various possible definitions were tested. For example, we originally constructed a one-dimensional definition that was not easily generalized to higher dimensions, so we had to revise said definition. As far as materials are concerned, the only materials used were textbooks and scholarly articles on related subjects.

Results
We were able to properly provide a new notion of a topological homeomorphism in terms of a dynamical system, namely a "lure". This definition is constructed for general n-dimensions. We are able to show that, with respect to the monoid of positive time under addition, a set M denoting the homotopy equivalence class of a manifold, and the "lure", this system satisfies the axioms of a dynamical system. However, we are able to generalize this notion beyond that of compact topological manifolds, to ordinary hyperreal fractal strings. That is, the natural hyperreal extension of an ordinary fractal string, the first appearance of such an object.

Conclusions/Discussion
In future work, we would like to apply the general idea of a "lure" to practical applications in economic systems, biological systems, and certain machine learning algorithms. Additionally, we would like to extend the notion of a lure to more general manifolds, as well as Iterated Function Systems

Summary Statement
We construct a new notion of a topological homemorphism in terms of a dynamical system.

Help Received
I'd like to thank Dr. Michael Maroun for his direct supervision and guidance, Dr. Michel Lapidus for his support, and Colin Aitken of MIT for his discussions on the subject.

CALIFORNIA STATE SCIENCE FAIR
2014 PROJECT SUMMARY

Name(s)

M. Evan Wildenhain

Project Number

S1430

Project Title

Supercolony: A Novel Ant Colony Optimization Algorithm for Solving the Traveling Salesman Problem

Abstract

Objectives/Goals

This study seeks to create a novel algorithm for producing superior solutions to the Traveling Salesman Problem by innovating upon Ant Colony Optimization techniques. A novel Ant Colony Optimization algorithm is developed that uses multiple, unique colonies in combination in order to produce superior solutions. It was hypothesized that the novel algorithm, "Multi-Colony System" (MCS), would perform 5% better than the "standard" ACO algorithm, Ant Colony System (ACS), at easy, moderate, and difficult TSP instances.

Methods/Materials

A quad-core Intel i5-3450 computer with 8 GB RAM was used for programming and running the experiment. Two variants of the novel algorithm, MCS, were coded by the researcher in Java; the researcher also implemented ACS for comparison purposes. MCS uses multiple colonies based on ACS with differing parameters to focus on exploration or exploitation. Each algorithm was run 10 times for 5040 seconds against TSP instances eil101, d198, pcb442, and pr1002.

Results

With the "easier" instances eil101 and d198, the two algorithms performed very similarly, with little difference in the mean tour lengths achieved by the algorithms. However, in the 442-city instance and 1002-city instance, the two variants of MCS outperformed ACS significantly, with Symmetric MCS outperforming ACS by as much as 20% in terms of mean tour length.

Conclusions/Discussion

ACS and MCS were comparable at "easier" instances with fewer cities, but MCS was a significant improvement when tested with the larger TSP instances of the study. More cities result in a greater number of possible solutions, which can increase the number of local maxima in the search space and thus the advantage MCS holds over ACS: diversity of tours. By pursuing multiple solutions at once, MCS can more efficiently avoid converging prematurely and search the solution space more quickly.

Summary Statement

This project develops and tests a novel Ant Colony Optimization algorithm for solving the Traveling Salesman Problem.

Help Received

Mother helped assemble board; Father advised researcher on program development.