CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Tristan J. Anderson

Project Number

S1401

Project Title

What Does Music Look Like?

Abstract

Objectives/Goals

The objective was to determine what different genres of music look and feel like.

Methods/Materials

Data was obtained by using a decibel meter to record the volume data and mathematical calculations using Excel to determine the beat frequencies per second. This data was then imported into Matlab (a multi-paradigm numerical computing environment) in order to conduct regression analysis through triangulation of the three parts of the data: time, frequency and decibels. These data sets were assigned to the x,y and z axes of a three dimensional graph. Time was assigned to the x-axis, frequency to the y-axis, and decibels to the z-axis. A 3D graphical model was then created of the surface structure for the array of points. Data was then converted to a .stl file in order to be accepted by a 3D printer. The models were then printed using a 3D printer.

Results

The finished plots both on screen and in tactile form produced non-uniform data models. The genre of classical was the most complex in that it had the most folds in the 3D surface. The genre of jazz and pop had much more simple structures and in turn smoother prints. My results enabled me to feel and see music in a new way that was perhaps never possible before. In the end I obtained my objective of seeing and feeling music.

Conclusions/Discussion

These results prove my hypothesis was correct in stating that classical music will have more complex structural surfaces. In the end I was able to attain my objective of seeing and feeling music. The information from this project furthers our knowledge into the subjects of not only computer science and programming but also into the physics of music and the quickly growing industry of 3D printing. The outcome of the models also suggests the potential complexity and emergent behavior found in music when modeled in space. The models from this project could potentially be applicable to people unable to hear who may not otherwise be able to hear music but could potentially learn to feel and see music. The results from this project could also be applicable in the architecture industry. To further my project I would like to reverse the outcome of my results by 3D scanning an object and composing my own piece of music with the data.

Summary Statement

This project focuses on creating a visual and tactile representation of music through three-dimensional data representation.

Help Received

None

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s) **Jennifer J. Choi**	**Project Number** **S1402**

Project Title

Quantum Random Number Generation by a Mobile Phone Camera Using Randomness of Photon Emission from a Light Source

Abstract

Objectives/Goals

Recently, security systems are based off of random number generators that do not truly generate ¡°random¡± numbers and are not completely secure. If random number generators are based off of random natural processes, especially those of quantum physics, they could produce better results. My objective was to analyze how quantum random number generation could improve this and in what aspects it was better than other forms of random number generation. I hypothesized that since photon emission is intrinsically random, quantum random number generators based off of photon emission should have much less bias and predictability in comparison to pseudo-random or hardware random number generators.

Methods/Materials

I held three different types of light – red laser, green laser, and white light – at a fixed distance from my Samsung Galaxy Note 3 camera and took pictures of each. The act of taking a picture ¡°measures¡± the location of photons and this creates random data since photons are emitted at completely random times. I then used ExifTool and other extractors to get binary data from the raw pixel data. The second part of my experiment was comparing the randomness of different types of random number generators (pseudo-random and hardware) with binary data I had for pictures from my phone. I created 1,000,000 digits of random numbers with each random number generator. I then analyzed data from all five different random number generators and found the distribution and standard deviation of the digits.

Results

All three of the quantum random number generators were the most ¡°random¡± and had the lowest standard deviations, while the pseudo-random number generator was the most predictable and had obvious bias.

Conclusions/Discussion

In conclusion, random numbers produced by photon emission processes were more truly ¡°random¡± and had less bias. Quantum random number generators should be implemented more in security systems in the future in order to create safer security systems.

Summary Statement

I explored how random number generation could be improved by using the natural randomness of photon emission to obtain quantum random numbers.

Help Received

Mom helped print and tape board

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s) Sanath K. Devalapurkar	**Project Number** **S1403**

Project Title

Preserving Algebraic Structures on Exact Quasicategories with the K-Theory Functor

Abstract

Objectives/Goals

The objective of this project is to establish the foundations for a new perspective on the algebraic K-theory of exact quasicategories, so as to allow for a simpler way to approach multiple problems involving the existence and preservation of algebraic structures.

Methods/Materials

The theory of quasicategories and quasioperads was used extensively. The classical perspectives on K-theory was used as a motivation for laying the foundation for a new perspective on the algebraic K-theory of exact quasicategories. The methods of simplicial homotopy theory and category theory were instrumental in proving our main results.

Results

The classical interpretation on the K-theory of an exact quasicategory is as a spectrum. My idea was to view these objects not as spectra, but rather as stable quasicategories. This shift in perspective allowed me to show that the K-theory functor respects module structures, and can be used to develop a homotopical derived Morita theory for algebras in quasioperads, which, in turn, allowed for the comparison of the K-theories of different exact quasicategories.

Conclusions/Discussion

My interpretation of K-theory as a stable quasicategory allows for multiple problems to be simplified. The main calculation shows that K-theory is a very good "algebraic functor" because it preserves module structures, thus, giving rise to a new derived Morita theory. This main calculation also has applications to a theory of quasi-n-operads, which I plan to explore in the near future.

Summary Statement

In this project, I describe, how viewing the K-theory of an exact quasicategory as a stable quasicategory (and not as a spectrum) can be beneficial to the study of homotopical algebra.

Help Received

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Alexandra S. Fischer

Project Number

S1404

Project Title

New Results on the Genus of Complete Graphs with Excised Edges and Parameterization of the Resulting Isomorphism Classes

Abstract

Objectives/Goals

This research project explores the effect on the genus of a complete graph Kv on v vertices when edges are removed and also explores how to classify the resulting graphs. The genus is an important characteristic of a graph, providing a single numerical measure of the complexity of the graph. Also of importance in graph theory is Kv, which is universal in the sense that all graphs are subgraphs of some Kv. Thus, understanding how the genus of Kv responds to the excision of edges is of major importance in graph theory.

Methods/Materials

Three major mathematical fields were used to study the genus of the graphs resulting from excising edges of Kv. Low dimensional topology was used to link the genus of a graph to surface theory, graph theory was used to establish bounds on the genus, and combinatorics were used to calculate the number of isomorphism classes of graphs obtained from Kv when edges are excised.

Results

Three major new results were established. The first considers how many edges can be removed from Kv without changing its genus. I was able to give a sharp answer to this problem by introducing a function F(v)= (v-2)(v-5)/2 mod 6. Then my first main result is that if F(v) or fewer edges are removed from Kv, then the genus of the resulting graph is the same as the genus of Kv. My second result is a general result, independent of the genus result, that parameterizes the isomorphism classes of graphs of the form Kv - h edges. This result states that these isomorphism classes are parameterized by the isomorphism classes of graphs with h edges, both connected and non-connected, and with no vertices of degree 0. My third result combines my first two results to parameterize the isomorphism classes of graphs of the form Kv - F(v) edges, which are then calculated. Thus this result explicitly classifies graphs derived from Kv by the removal of F(v) edges. By my first result, these graphs then have the same genus as Kv. Applications of these results to complex networks, such as social networks and the internet, are also given.

Conclusions/Discussion

This project gives new results in graph theory related to the stability of the genus of Kv with respect to the excision of edges. The parameterization of the isomorphism classes of the resulting graphs is calculated. These results can be extended in two directions, to other well-known classes of graphs and to the excision of edges that decrease the genus of Kv.

Summary Statement

This research project presents new results on the stability of the genus of a complete graph Kv with respect to the excision of edges and gives an explicit parameterization of the resulting isomorphism classes of graphs of the form Kv-F(v).

Help Received

My mother helped me build the various physical models that I used to illustrate my results.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)	**Project Number**
Aratrika Ghatak	S1406

Project Title

A Low Cost Traffic Signal Preemption Solution for Emergency Vehicles Leveraging Mobile and Cloud Computing Technologies

Abstract

Objectives/Goals

Emergency vehicles on duty are generally allowed to pass a red light. However that is not always be possible if there are several cars blocking its path. My project objective is to leverage mobile and cloud computing technologies to create a low cost system that will be able to turn the traffic light green right before it reaches the traffic signal. This will ensure that there is no delay in responding to an emergency situation and will prevent potential road hazard.

Methods/Materials

I have created two mobile apps using MIT AppInventor2 for Android. The first app (named as V2S Communication) will be with emergency vehicle (EV) driver#s android device. When destination is entered, it will determine the route and traffic signal data on its path. Traffic signal data is generally available with the city administration. For my project, however, I have manually entered it into a database. When the driver requests to synchronize traffic lights, the app will calculate estimated time of arrival (ETA) to each cross road using time/distance formula and will store such information into a Fusion Table on Google cloud.

The second app (named as Traffic Light Simulator) simulates how traffic signal will work in each cross road. It will retrieve ETA of the emergency vehicle (EV) to a specific cross road from Fusion table on Google cloud. It compares current time to the ETA and turns traffic light to green ten seconds prior so that the EV can smoothly pass the cross road.

I used free MIT AppInventor2 as app development tool and Google Fusion Database as on-cloud database to store information.

Results

My first application demonstrated the ability to calculate ETA to each traffic light and then stored the information into on-cloud database. The second app simulated how each traffic light will retrieve ETA from Google cloud and will change it to green ten seconds prior to the ETA, as expected.

Conclusions/Discussion

My solution has proved to be a low-cost and efficient way of solving the traffic preemption problem for emergency vehicles. Although there are other traffic system preemption devices already in use, their adoption rate isn#t too high. Most of them are too expensive or use sensors which can be interfered easily. The price of my solution will include the minimal cost of wireless connectivity. It also demonstrates how cloud and mobile technologies can work together for faster emergency response to save life.

Summary Statement

My project provides a low cost solution for faster emergency response via Vehicle to Signal (V2S) Communication leveraging state of the art Mobile and Cloud technologies.

Help Received

Thanks to my teacher, Mr. Doug Miller to recommend MIT AppInventor2 to develop the mobile apps; Thanks to my dad to investigate and explain the existing preemption solutions; Thanks to my brother to share his android tablet to test my mobile apps.

Name(s)

Ram S. Goli

Project Number

S1407

Project Title

Merging Machine Learning and Archaeology: Using an Artificial Neural Network to Predict the Implementation of Disk Beads

Abstract

Objectives/Goals

My objective is to create a computer algorithm that can accurately predict the implementation of a Clam Shell Disk Bead (CSDB) by recognizing pattens in predetermined beads, mainly patterns in the beads' diameter, thickness, and aperture. The algorithm should be able to predict whether a bead of unknown purpose is either a money bead or a basket bead.

Methods/Materials

I created an ANN in the Octave programming environment, and trained the network with data provided by archaeologist Katherine Dixon. I performed various network optimizing tasks such as plotting learning curves and validation curves, and testing different network structures to ensure that the network has the highest possible prediction accuracy on test data beads.

Results

The ANN I developed accurately predicts the implementation of %75.76 of beads given to the network (money bead or basket bead) by checking whether the implementation predicted by the network matched the predetermined implementation.

Conclusions/Discussion

It is possible to model CSDBs and predict their implementation with an Artificial Neural Network, but not with tremendously high accuracy. Due to many of the predetermined money and basket beads having similar measurements, the prediction accuracy of the network could not be substantially increased. However, my work will hopefully allow archaeologists to easily predict functions of beads collected at future digs, and lead to further bridging the fields of machine learning and archaeology.

Summary Statement

In this project, I developed an artificial neural network to predict the implementation of clam shell disk beads found at archaeological sites.

Help Received

Archaeologist Katherine Dixon provided bead data

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Ajia Grant

Project Number

S1408

Project Title

Can You Beat the Market?

Abstract

Objectives/Goals

My objective was to create a trading strategy to achieve a higher rate of return than the overall stock market. I predicted that it is possible to outperform the buy-and-hold investment strategy by using technical analysis based on historical price and volume data to determine buying and selling decisions. I then implemented a genetic algorithm to find successful strategies using combinations of analyses.

Methods/Materials

The algorithm was trained and evaluated on the SPY index fund, using data from 1995 to 2004 and 2005 to 2014, respectively. Trading decisions were evaluated on a daily basis, and return on investment was used to rank the strategies in each generation. I varied population size, number of generations, and trading cost imposed during training to analyze the algorithm's sensitivity to these parameters. I repeated the experiment ten times to verify the results were reproducible.

Results

The algorithm was usually able to beat the buy-and-hold strategy, sometimes by a lot, but not every time. The mid-sized population produced the best results. Increasing the number of generations selected increasingly complex strategies, which, beyond some point, did not consistently improve results, especially for small population sizes. Complex strategies that performed spectacularly well (up to 10x return on investment) on the training period did not do well on the evaluation period. Significant trading costs reduced trading frequency and improved performance compared to insignificant costs.

Conclusions/Discussion

I concluded that it is possible to outperform the buy-and-hold method using technical analysis, at least in the case that there is a significant market downturn, as in the periods analyzed. While genetic algorithms can reliably identify profitable combinations of analyses, overfitting by "getting lucky" on the training data is a significant problem. Increased trading costs somewhat mitigated overfitting by penalizing excessively frequent trading.

Summary Statement

I developed a computer program to make automated trading decisions based on analyses of historical price and volume data that achieved superior return (in simulation) than simply buying and holding the SPY index fund.

Help Received

Father helped find relevant research, reviewed the code, and scripted the data collection.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

George Hou

Project Number

S1409

Project Title

Separating Mixed Signals in a Noisy Environment Using Global Optimization

Abstract

Objectives/Goals

The main objective of my research is to design improved blind source separation methods guided by rigorous error analysis. Blind source separation aims at recovering source signals from mixed signals without detailed knowledge of the source signals or the mixing process. Currently there is no rigorous error analysis of blind source separation methods, and existing methods are known to be ineffective in separating source signals in a noisy environment. My research provides precise solvability conditions under which blind source separation is guaranteed to give accurate signal recoveries in a noisy environment.

Methods/Materials

I propose a new global optimization method that minimizes the cross-correlation of the recovered signals, which is key to proving an exact recovery of source signals. And, unlike current methods that attempt to alleviate background noise by using de-noising methods, I treat the noise signal as a separate source signal and obtain an extra recording of the mixed signals. In the case when the mixing matrix is ill-conditioned, I use QR factorization to pre-process the mixtures to obtain a well-conditioned mixing matrix, which produces more accurate recoveries.

Results

I carry out rigorous stability analysis for my global optimization method and give precise solvability conditions that guarantee an accurate recovery of source signals. To the best of my knowledge, this is the first error analysis of blind source separation methods. Repeated numerical experiments suggest that my method is more robust than existing methods and provides more accurate signal recoveries.

Conclusions/Discussion

Given the widespread applications of blind source separation, my optimization method and my error analysis can play a substantial role in this emerging field. My rigorous mathematical analysis, which can be applied to designing more robust noise reduction algorithms, provides a solid foundation to this field and demonstrates the importance of mathematical analysis in engineering and scientific applications. In particular, my research has the potential of improving the current design of hearing aids.

Summary Statement

Through providing precise solvability conditions and error analysis of blind source separation methods, I have developed a robust method that successfully separates mixed signals into their respective source signals in a noisy environment.

Help Received

Professor Jack Xin supervised the progression of my research and provided useful discussions.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s) **Abe N. Jellinek**	**Project Number** **S1410**

Project Title

Lingua: A Programming Language for the Uninitiated

Abstract

Objectives/Goals

The purpose of this computer science engineering project was to create a programming language usable by anyone, from a total beginner to an advanced programmer. Therefore, it was designed to be approachable and simple, yet encompassing of a wide variety of programming paradigms and concepts. Its syntax was based on languages such as Java and C, but simplified and refined in many ways in order to be approachable by new students as well as experienced programmers. I called it Lingua, after the Latin word for language.

Methods/Materials

Materials:
- Java 8 - a cross-platform programming language
- Google Guava - Google#s standard library of collections for Java
- JUnit 3 - an automated unit test framework

I researched parsing algorithms and settled on Pratt parsing, a variant of recursive descent parsing. It makes operator precedence quite easy, and that was a huge help.
Next, I developed a simple parser for mathematical expressions, which then gained support for variables, functions, data types, looping, and many other essential parts of modern programming languages.
Finally, I wrote a tree-walk interpreter for the language, because, while not the fastest method, it#s quick to prototype. It works by evaluating each top-level expression, which then evaluates each of its child expressions, and so on, until it reaches the #leaves# of the tree.

Results

The entire implementation came out to nearly 10,000 lines of code in 93 classes. Much of that is the lexer/parser (the part of the implementation that cuts text into tokens, and then joins tokens into expressions), which is relatively complex (not hard to understand, but large). The interpreter is surprisingly simple, and performance is much, much better than you would expect from a tree-walk interpreter. The Java VM is very smart with its optimization, and successive runs get significantly faster as it optimizes and JIT-compiles the code.

Conclusions/Discussion

I accomplished my goal in its entirety. The language is approachable by beginners, has a fairly large feature set, and is quite flexible. It is fast and complete enough that I actually believe that it could be used as a teaching tool, and all it needs to really succeed is a larger standard library and more documentation.

Summary Statement

A programming language with a focus on usability by beginners and a broad feature-set.

Help Received

I received very small amounts of help with my board from my family.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Tanisha Joshi

Project Number

S1411

Project Title

The 99¢ Clinical Trial: Accelerating Trials in Software for ErbB2 Pathways and Lapatinib on Metastatic Breast Cancer

Abstract

Objectives/Goals

6540 breast cancer clinical trials are being performed in the world and the drug life cycle takes an average of 14 years. The average pre-tax industry cost per new prescription drug approval (inclusive of failures and capital costs) is $2.55b. Accelerating fast failures can dramatically reduce drug approval costs and improve drug behavior predictability substantially. With this model/software, pharmaceutical scientists can predict drug behavior in 21 min (instead of 21 days) by simulating virtual trials which are predictive pharmacogenomic models of ErbB2 activated signaling pathways in conjunction with Lapatinib (CAS 388082-78-8). Accelerating the experiment (1440 times faster) involves modeling the Mechanism of Action of Lapatinib ditosylate and Capecitabine in a cell region physiology exhibiting overexpression of ErbB2 (2-6 copies), a sub-cellular biomarker of advanced (Stage IV) metastatic breast cancer. The goal is to successfully complete a virtual trial in software to predict the drug response of Lapatinib for HER2-positive advanced metastatic breast cancer, in a digital micro-biopsy cell region matrix. The results will be validated with a corresponding clinical trial on women in China.

Methods/Materials

MacBook Air 4GB 1600 mHz Intel Core i5 was used for testing.

Results

It was observed that the natural DNA repair mechanisms remained unaffected by tumor growth, because the dysregulation of the Ras mediated MAPK signaling pathway only amplified cell proliferation signals. The Clinical Benefit Rate (CBR) was found to be in the range of [43.2%, 71.3%]. Also the adverse drug reaction (ADR) was in the proximity of 3.8% (at 5% level of significance) Lapatinib molecules were consistently engaged in a repair mechanism that increased their utilization and left a smaller quantity of substrate.

Conclusions/Discussion

I developed a clinical drug study tool, that follows a computationally and economically scalable (less than $1 per patient) model which simulates the ErbB2 signaling pathways and their response to Lapatinib.The software algorithms that I built would potentially allow pharmaceutical scientists to rapidly permutate over millions of patient biologies and candidate drug structures based on rule set specificity. All software is on GitHub and available upon request.

Summary Statement

The project involves pharmacological modeling of a advanced/metastatic breast cancer clinical trial in software based on measurement of drug response on pharmacogenomically generated patient biologies.

Help Received

Industry Drug Researcher Archana Gangakhedkar at Xenoport Clinically Validated Software Results

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Sarah H. Kazmie

Project Number

S1412

Project Title

DermatoScan: Machine Vision, Analysis, Learning & Natural Computing Optimizations for the Early Detection of Skin Cancer

Abstract

Objectives/Goals

A dermatoscope is a handheld device used by doctors to view skin tumors. Very few dermatoscopes provide any form of intelligent analysis or diagnostic support. Machine vision, image processing, machine learning and artificial intelligence could be combined to develop and train a fuzzy-logic neuro-network to identify pathological high-risk skin condition. Such a system could be embedded in an intelligent device, creating an easy to use and inexpensive dermatoscope that could be sent home with patients to assist them in tracking their own skin conditions. This software could also be integrated into and distributed in the form of an app for a computer, tablet or smartphone.

Methods/Materials

Developed a C++ program using the OpenCV library to capture and isolate images of skin lesions. Created numerical methods to extract and quantify visual features, including brightness, contrast, symmetry, concavity and color variability. Collected a library of sample images, and implemented a simple fuzzy-logic neuro-network simulation function, which analyzes the measured features to assess the risk of malignancy. I applied this function as a fitness test to implement a genetic learning algorithm, which tested and adjusted the weighting parameters applied to each feature. I tested, adjusted, optimized and repeatedly re-tested these algorithms.

Results

I tested the genetic training algorithm many times. Even after a series of 950 generations, the success rate never surpassed 89% Adding images to the training set produced a consistently higher success rate of 92%, but changing the internal transfer/fitness function to expand the sensitivity of some variables and then tightening the cutoff criteria for determining fitness, produced a very significant improvement, consistently achieving 100% accuracy rates within the known sample set.

Conclusions/Discussion

The results support the hypothesis. The evolutionary genetic learning process quickly and effectively adapts its weights and parameters to achieve a very high sensitivity and specificity in the recognition and classification of skin cancer lesions. Given a broader sample set or a wider range of analyzed visual features the effectiveness and efficiency might further improve. These algorithms could be embedded into a variety of real-time micro-controllers and mobile devices.

Summary Statement

Computer vision, image analysis, feature extraction and a genetic machine learning process can evolve and determine an effective set of weights and parameters for a fuzzy-logic neuro-network to recognize and evaluate skin cancer.

Help Received

I#d like to thank Harry Evry for introducing me to OpenCV and for suggesting I learn about fuzzy-logic and natural computing. Thanks also to Dr. Malhotra for helping edit, organize and improve my research report and project documentation.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Kun Woo Kevin Kim

Project Number

S1413

Project Title

Combining Programming Languages to Improve Performance

Abstract

Objectives/Goals

The goal is to generally improve the performance of any device through combining multiple programming languages.

Methods/Materials

The important Materials are:
Computer
Eclipse Java Software
Eclipse C++ Software
The Procedure Is:
1.Launch the Eclipse program for the Java environment
2.Enter in the code for multiplying matrices and a way to time it
3.The result is given in milliseconds, record it.
4.Repeat step 2-3 20 times.
5.Launch the Eclipse program for the C++ environment.
6.Enter in the code for multiplying matrices and a way to time it
7.The result is given in milliseconds, record it.
8.Repeat step 6-7 20 times.
9.Launch the Eclipse program for the Java environment once again
10.Enter in the code multiplying matrices utilizing the JNI and a way to time it
11.The result is given in milliseconds, record it.
12.Repeat step 10-11 20 times.
13.Analyse Collected Data.

Results

The results were that through the insertion of JNI, the performance indeed has been increased by 85%.

Conclusions/Discussion

The results are all pointing towards the fact that indeed JNI helps improve the performance of coding with the computers. Also, the times that this took seems to be fluctuate a bit between trials but still maintains a similar zone of time with repeated trials. By averaging the 20 trials, I believe that I got rid of the fluctuation as much as possible. The efficacy of the JNI has now been proven to be useful even in computers with this experiment. Also, one of the major parts being the fact that the performance increased by 85 percent is quite amazing with the simplest insertion of the JNI protocol.

Summary Statement

Combining Java and C++ o see if the performance improves.

Help Received

None-except the internet.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s) **Simon L. Kuang**	**Project Number** **S1414**

Project Title

Predicting Chromosome Conformation from Epigenetic Features

Abstract

Objectives/Goals

The objective is to predict the results of 40kb HinDIII Hi-C experimentation on the mouse ESC genome from 21 documented mouse epigenetic features.

Methods/Materials

Around 200 million 22-dimensional datapoints were drawn from the NIH histone database and from the Hi-C paper. Characteristics of the data required development of a novel machine learning meta-algorithm using inverse probability Metropolis-Hastings sampling, bagging with 100 independent samples, single-layer perceptrons optimized using Levenberg-Marquardt training, and intelligent weighting based on in-sample prediction correlation. Chromosome 1 was used for training, chromosome 2 for validation, and chromosomes 3 to X for testing.

Results

The model's predictions scored a Pearson correlation coefficient between 0.7 and 0.8 on all chromosomes. In-sample RMSE and out-of-sample RMSE decreased on all predictors over the 5 training epochs, and in-sample correlation and out-of-sample correlation increased on all predictors over the 5 training epochs.

Conclusions/Discussion

My method performs extremely well on this unusual dataset. In the next decade, spatial distance will be more insightful, cheaper and faster to compute, and easier to analyze in clinical situations than the DNA sequence itself.

Summary Statement

A novel machine learning meta-algorithm is necessary and effective for studying the 3-dimensional conformation of the mammalian genome.

Help Received

Used computing facilities at PSU Center for Comparative Genomics and Bioinformatics

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s) Frank Liu	**Project Number** **S1415**

Project Title

Are All Irrational Numbers Equally Irrational?

Abstract

Objectives/Goals

The purpose of my project is to investigate whether or not all irrational numbers contain patterns within their continued fraction form. Additionally, if they do have patterns, are there patterns amongst the patterns themselves?

Methods/Materials

1. Choose irrational numbers to test. I chose the irrational numbers pi, e, the square roots of 2 through 100 (excluding the perfect squares in between) and the cube roots of 2, 3, and 4.

2. Next, express these irrational numbers in continued fraction form.

To do this, start by entering the irrational numbers in a calculator. The calculator will give a trailing decimal approximation.

Subtract the integer portion from the number, and record that integer as the first data value. For example, in (approximately 3.14159#), subtract 3 from the calculator#s approximation to get .14159...

Next, perform 1 divided by the result. Continuing with our example, we do 1 divided by .14159# to get 7.06251#

3. Repeat step 2 at least 9 more times and record the results. In our example, the next integer portion to subtract and record is 7.

Record all of the results calculated in our list. In our pi example, the list would look like:

3,7,15...

4. Look for patterns in and amongst the lists formed.

Results

Expressing irrational numbers in continued fraction format does indeed reveal startling patterns. For example, the square root of 17, which seems to be a nondescript, endless number (4.123105626#) in decimal format reveals itself to be quite amazing in continued fraction format with its simple pattern of all 8s. Even more, it appears that every square root has a pattern within it when expressed in continued fraction format. Nevertheless, it is clear there are some other irrational numbers that still don#t have patterns to them- such as or the cube roots of 2-4.

Conclusions/Discussion

My hypothesis was partially correct. In my hypothesis, I thought that all irrational numbers still wouldn#t exhibit any patterns in continued fraction form. This was true for some irrational numbers, such as and the cube roots of 2-4, but was untrue for all of the square roots that I tested. Even more, I discovered some startling patterns inherent amongst the patterns of the square roots themselves.

Summary Statement

Expressing irrational numbers in continued fraction format can reveal simple patterns.

Help Received

Father helped show how to write numbers in continued fraction format

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Yelena Mandelshtam

Project Number

S1416

Project Title

Arrangements of Minors in the Totally Positive Grassmannian and Sturmfels' Triangulation

Abstract

Objectives/Goals

We discuss arrangements of equal minors in the totally positive Grassmannian. It was previously shown that arrangements of equal minors of largest value correspond to the simplices in the Sturmfels' triangulation. Here we investigate a major problem of arrangements of equal minors of m-largest value and find relationships between the minors and maximal simplices of certain cubical distance in the Sturmfels' triangulation.

Methods/Materials

To conduct this research, I used several mathematical tools. I used the combinatorial theory of sorted sets, Skandera inequalities, the theory of the Totally Positive Grassmannian, the graphs of triangulations, and general combinatorics and linear algebra knowledge.

Results

Several major results were obtained through this research. I defined two new notions- one was the notion of movements between maximal minor arrangements, and the other was the notion of cubical distance, which says that two vertices in the graph of the Sturmfels' triangulation are of distance one if they lie on the same hypercube of any dimension. I formulated and proved two theorems about the arrangements of second largest and third largest minors. Then I posed a conjecture for the general case which I proved in several subcases.

Conclusions/Discussion

I found and proved important properties and relationships between arrangements of minors in totally positive matrices and other important combinatorial objects. Besides completely describing the arrangements of minors of second and third largest value, I formed a conjecture for the general r-th largest minors in terms of the new notion, multidimensional cubical distance, that I introduced, and showed that the conjecture is true in several subcases. It seems that the rest of the conjecture may be within reach using methods similar to those used in my paper. The surprising relationships I found between totally positive matrices and the Sturmfels' triangulation could pave the way for future advances in physics, computer science, and mathematics. A paper containing these results has been accepted to FPSAC 2015, the premier combinatorial conference.

Summary Statement

I studied arrangements of equal minors in the Totally Positive Grassmannian, a combinatorial object which could provide insights into physics, computer science, and mathematics.

Help Received

Research mentor and advisor: Miriam Farber, graduate student at MIT

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Clarence M. Nakano

Project Number

S1417

Project Title

Biological Electron-Transfer Dynamics in Multiheme Cytochrome Complexes

Abstract

Objectives/Goals

Electron transfer (ET) governs all known energy-conversion processes in biology. A remarkable example is the recent discovery of rapid ET along electrically conducting bacterial nanowires produced by the bacteria Shewanella oneidensis MR-1. The outer-membrane cytochrome, MtrF and OmcA, are hypothesized media for ET, but how these multiheme cytochromes are assembled into a conducting complex remains a mystery. Thus, the goal of my project was to determine the structure of the MtrF-OmcA complex and visualize ET dynamics in it to better understand the underlying electric conduction mechanisms.

Methods/Materials

Here, I determine the structure of the MtrF-OmcA complex and study ET dynamics in it by combining the use of a homology modeling server, protein docking software, kinetic Monte Carlo (KMC) simulation, and visualization. I developed a computational framework including the entire workflow. In particular, I developed a C-RANK program to screen complexes according to biological plausibility as well as a plugin to the Visual Molecular Dynamics software named ETViz to animate ET dynamics.

Results

My visual simulation results reveal novel nonequilibrium phase transitions with which Shewanella efficiently responds to a change in its electrochemical environment. The KMC results, when compared with experimentally observed respiration rates, suggest that life operates around the triple phase junction.

Conclusions/Discussion

My simulation and visualization results shed useful light on boosting the efficiency of Shewanella-based microbial fuel cells by increasing the ET rate toward solving the global energy problem. Currently, I am working on a larger study of ET mechanisms along a lattice of outer-membrane cytochrome complexes on the entire bacterial nanowire.

Summary Statement

I developed an electron transfer visualization software to reveal novel nonequilibrium phase transitions with which a bacterium efficiently responds to a change in its electrochemical environment.

Help Received

Research Supervised by Prof. Mohamed Y. El-Naggar (USC) and Prof. Tao Wei (Lamar University)

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Adam Z. Noworolski

Project Number

S1418

Project Title

Games vs. Work: Analyzing Keystroke Patterns to Detect Specific Activities on a Computer

Abstract

Objectives/Goals

My goal of this project was to accurately show how long I was on the computer, and what I was doing, based on mouse movement, keystrokes, and CPU usage.

Methods/Materials

The only necessary materials for my project include my computer, my code for analyzing and logging keystrokes, and Keen.io cloud-based storage system.

First, created code to log keystrokes to a file. I tested the software by pressing a specified sort of keys and looking at the output. Next, I got keypress patterns from actual activities. To accomplish this, I ran the software when I did a specific activity, and kept a note of the start and end time. When the activity ended, I looked at the logged key file, and wrote down some of the various patterns that came up. I continued taking more data from different activities. After I knew which specific patterns relate to which activities, I wrote a program to detect when those specific patterns occur, and therefore detect when each activity is being played, and a lack of patterns would lead to mean that the user is not doing one of the specific activities. In addition, I have the code log how long each activity lasts, and comparing the actual numbers to the program#s prediction can calculate the accuracy of the program.

Results

I was able to show how my program worked by testing it on League of Legends games that I played. My code had a greater than 76% accuracy of detecting the correct time, and any error was within five minutes of the correct time. Because this program was not as accurate as I would like, I will continue to search for the #perfect pattern# where it is evident ONLY at the start of every activity, and never false alarms.

Conclusions/Discussion

Later, I would like to add mouse and CPU usage comparison to increasingly be able to identify what activity is happening on the computer, so that it can have more accurate results.

In the future, I would like to have more patterns, as of right now, I only have a few patterns, whereas, in the best scenario I would have many patterns for activities so that I could have as much accuracy as possible. In addition, I have developed a large interest in Machine Learning and Neural Networks recently, and I hope to implement them soon to find new patterns.

Summary Statement

Use programs and software to detect keystrokes to understand what activity is being done on the computer, and how long it lasts.

Help Received

Parents helped edit and glue slides on poster

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)	**Project Number**
Charles S. Noyes	**S1419**

Project Title

BitAV: Fast Decentralized Anti-Malware by Distributed Blockchain Consensus and Feedforward Scanning

Abstract

Objectives/Goals

Anti-malware software is responsible for the protection of some of our society#s most important infrastructure. I have designed and implemented a novel anti-malware system (BitAV) that performs considerably better than all industry solutions I was able to test. Through the use of novel consensus protocols (adapted from those defined in the Bitcoin whitepaper) and fast, bit-vector-based scanning mechanisms I was able to greatly increase the speed of update propagation and malicious file identification.

Methods/Materials

Networking software was written in Python and the scanning mechanism in C++. Blockchains were chosen for their ability to be shared across a P2P network with assurances of canonicity and their ability to store arbitrary (but cryptographically verifiable) data. Bloom filters (and their derivatives) were chosen as data structures to be used in the feedfoward bloom-bloomier scanning mechanism for their speed in probabilistic pattern matching.

Results

Two tests were performed, one that measured propagation time of new malware identifiers and one that measured time taken to scan different malware samples. Scanning time showed the largest gain in performance, with an average increase of over 1,400% in speed ($p < 0.0001$, three sigma confidence), increasing by a factor of 2 to 3 with realistic input file-sets. Finally, the identification and propagation of novel malware variants happened, on average, 500% faster ($p < 0.00001$, four sigma confidence). All tests were conducted against (between 10 and 46) industry standard and OSS solutions.

Conclusions/Discussion

Ultimately I accomplished what I set out to achieve with this project. Further applications include over-the-air malware filtering and decentralized STIX/MAEC-over-TAXII networks, both of which are made possible with this software architecture. Additionally, my novel blockchain variant can be easily adapted to work with distributed associative memory networks, which make things like blockchain-resident neural networks, decentralized prediction markets, etc. a very real possibility.

Summary Statement

I created an application that efficiently protects users against malicious pieces of software, which works trustlessly across a peer-to-peer network using protocols similar to Bitcoin's.

Help Received

VirusTotal allowed me to use their databases of known malware and industry standard anti-virus solutions to compare the performance of my solution against currently available competitors. Chi So, professor of information security at USC, provided useful, tangentially related, discussion but was not involved in any

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Peizhu Qian

Project Number

S1420

Project Title

Angle Analysis of Multiple-Line Intersections

Abstract

Objectives/Goals

(1) To analyze the intersection of 'n' non-parallel lines locating on the same plane, and try to find: A) How many angles (less than 180 degree) that 'n' lines can form; B) What is the value of the smallest angle(s); C) How many right angles, acute angles, and obtuse angles that 'n' lines can be formed.
(2) To extend the results that obtained in part (1) to the intersection of planes and the intersection of circles.

Methods/Materials

Begin with the simplest case: the intersection of 2 lines. Summarize the properties of the 2-line intersection including the number of intersecting points, the number of angles, the value of the smallest angle(s), and other findings. Then, gradually add more lines, one by one. Make the new line intersect with all previous lines respectively. Record the properties. When the number of lines is greater than 2, there are many possible ways for the lines to intersect. Now, rearrange the lines so that they intersect in as many ways as possible. Record the number of intersecting points and the number of angles that the lines form in each different case.

Results

When the number of lines 'n' is fixed: 1.the number of angles is always the same regardless of the number of intersecting points. 2.the value of the smallest angle(s) is also the same regardless of the number of intersecting points. 3.the number of acute angles is always the same as the number of obtuse angles. 4.the maximum number of intersecting points is the same as intersecting axes that formed by 'n' planes.

Conclusions/Discussion

I. Intersection of lines:
1)Find the maximum number of intersecting points when n lines intersect at different points, formula: n(n-1)/2
2)Find the number of angles when n lines intersect, formula: n(n-1)
3)The value of the smallest angle, formula:less or equal to 180 degree/n
4) The number of different types of angles
When there are two lines perpendicular:
Right angles: 4; Acute angles=Obtuse angles=n(n-1)-2
When no perpendicular lines:
Right angles: 0; Acute angles=Obtuse angles=n(n-1)
II.Intersection of Planes

Summary Statement

When there are 'n' lines intersecting on the same plane, the number of angles, the value of the smallest angle(s), and the number of right angles, acute angles, and obtuse angles can be determined by the formulas found in this project.

Help Received

1.My math teacher, Juan Gracia, and my grandma helped me go over my summary to make sure that it is understandable. 2.A math professor from Mendocino College, Deborah H. White, came to talk to me about further study of my project after the county science fair. 3.My classmates lent me glue and colored

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s) Kalyani Ramadurgam	**Project Number** S1421

Project Title

Face Recognition via Projection into Lower Dimensions Using Ensembles of Fisherfaces and Eigenfaces

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, hidden, or under extreme lighting. Novel recognition of hidden faces is urgently needed in settings of both military security and personal use. Using new methods of dimensionality reduction to significantly increase the density of data and focus on only orthogonal features, 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 some preprocessing steps, as well as Scikits for elements of dimensionality reduction. I processed and filtered pictures in the dataset with Gaussian blurring, high-pas contrasting, and greyscale conversion. To crop the faces, I used a derivative of the rejection cascade of weak classifiers. Principal Component Analysis, Fisher's Linear Discriminant, and pixel patterns were all used simultaneously with individual parameters. After synthesizing individual inferences, I used a novel confidence generating algorithm coupled with a new consensus algorithm to generate a final inference.

Results

Using the complete CMU dataset with both obscured and unobscured faces, the system generated an accuracy of 95.2%, which is comparable to modern use. With only the subset of frontal faces, an accuracy close to 100% was generated. When faced with only obscured faces, the algorithm had an accuracy of 67.5%. Current techniques have proven to be almost completely random when only given hidden faces, which is about a 4.95% accuracy. So, the algorithm presented in this project provides a significant improvement when working with hidden faces.

Conclusions/Discussion

I have successfully created a face recognition system that recognizes obscured faces with a higher accuracy than current techniques. The combination of preprocessing steps and dimensionality reduction synthesis 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 place.

Summary Statement

By combining processing and new dimensionality reduction, I created a powerful facial recognition system that accurately recognizes hidden faces under extreme lighting, 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
2015 PROJECT SUMMARY

Name(s)

Amrit Rau

Project Number

S1422

Project Title

Lightweight Path-Inferential Optic Flow Collision Detection for Mobile Robots

Abstract

Objectives/Goals

In this investigation, a lightweight, optic flow-based collision detection algorithm capable of running in real-time was developed with the goal of achieving key benchmarks for both safety and mobility (100% and 80% respectively) while running in real time aboard a mobile robot. Drawing on gyroscopic data, the detector calculates an inferred path triangle, enabling it to respond appropriately to real-world situations in which the camera heading vector is not necessarily in alignment with the vehicle#s heading vector, such as rotation and slip.

Methods/Materials

The detector was implemented aboard a HP Envy TS 14 PC in Python 2.7 using the NumPy and OpenCV libraries. A LEGO Mindstorms NXT 2.0 kit was used to construct a mobile robotic platform with a turntable upon which an iPhone 4s was mounted. The iPhone streamed Grayscale video to the PC for processing. This detector platform was placed in a testing environment, which consisted of a level testing surface perpendicular to a video monitor.

Static and dynamic virtual obstacles were displayed on the screen. By rotating the iPhone mounting turntable so that the perceived origin of normal progressive optic flow vectors was not at the center of the camera#s field of view, conditions of rotation and slip were simulated. A true positive (TP), true negative (TN), false positive (FP), or false negative (FN) was recorded for each trial; the safety and mobility scores of the platform were calculated using the method of Chalupka, et al.

Results

Out of the 64 trials, 100% net safety and 86.5% net mobility were achieved. In other words, the algorithm never failed to detect impending collisions and incorrectly detected an impending collision approximately 15% of the time.

Conclusions/Discussion

These results constitute evidence in support of the safety and mobility hypotheses; the lightweight, path-inferential optic-flow based collision detector developed in this investigation did serve as an effective cue to collision. In the context of other similarly lightweight collision detection algorithms, the detector developed in this investigation is highly successful. My algorithm has a plethora of applications, from assisting the navigation of the blind or visually impaired to refining emergency response robotics to mitigating mission risks for orbital spacecraft.

Summary Statement

A novel lightweight, optic flow-based collision detection algorithm capable of running in real-time was developed with the goal of achieving key benchmarks for both safety and mobility.

Help Received

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Manan A. Shah

Project Number

S1423

Project Title

Improving the Accuracy of Sentiment Classification: A Novel Synthesis of Computational and Analytical Methods

Abstract

Objectives/Goals

The determination of individuals' mood in a review of a restaurant or the public sentiment of a political campaign are incredibly important statistics that cannot be accurately estimated manually or with simple computational techniques.

The purpose of this project is to a) make use of supervised and weakly supervised machine learning algorithms to accurately classify the sentiment of a string of text, and b) incorporate negation handling, word n-grams, feature selection by mutual information, subjectivity classification, and polarity determination to improve classification of sentences. This project is unique as it is the first in the field to holistically explore a novel combination of both supervised and weakly-supervised machine learning models.

Methods/Materials

The two approaches studied were tested against corpora of data from IMDb, Amazon reviews, and Twitter for accuracy, precision, and recall. For each specified dataset, numerous iterations were run with different sample sizes ranging from 15 to 100. The primary analysis involved the use of IMDb pre-classified polar movie reviews. Every review was split into sentences, which were preprocessed, classified for subjectivity and polarity, and stored for future predictions.

Results

After training, feature selection by mutual information, and further textual analysis, the supervised model yielded an average accuracy of 88.7%. The weakly supervised model predictions continually increased in accuracy and were able to consistently predict results with an accuracy of greater than 83% after only 600 iterations. The weakly supervised model was more adept at making predictions on novel data due to its use of pattern matching and objectivity classification, whereas the supervised model prevailed at classifying sentences similar to its training set.

Conclusions/Discussion

The weakly supervised model improved on the foundations of the supervised model. The addition of subjectivity and polarity classification as well as feature selection vastly improved accuracies as only highly subjective sentences were included in overall calculations. The primary difference between the supervised and weakly supervised model was the analysis of linguistic patterns in sentences, which allowed for better classification of unseen cases.

Summary Statement

This project compared and improved weakly supervised and supervised machine learning models using linguistic analysis, polarity and subjectivity classification, and negation handling to effectively classify the sentiment of provided text.

Help Received

Parents helped with the board assembly. Computer Science teacher and mentor Dr. Eric Nelson helped with algorithm testing.

Name(s)

Arjun S. Subramaniam

Project Number

S1424

Project Title

Think-To-Type: Recognizing Letters from EEG Patterns Using Machine Learning

Abstract

Objectives/Goals

The objective of this project was to determine a proof-of-concept in recognizing letters that subjects are thinking about based on their brain's EEG patterns. Our goal was to achieve satisfactory accuracy (>= 70%) in classifying three letters - A, B, and C, from a subject's brain waves.

Methods/Materials

Materials included the Biopac EEG Machine, which consisted of three electrodes attached to the scalp, and the accompanying software. In addition, the WEKA machine learning software was used for training and testing the algorithms. We collected data from three subjects and had each one sound out the desired letter in their head at regular intervals for a period of 10 seconds. After initial pre-processing, we performed a Fast Fourier Transform and ran the FFT data through several machine learning algorithms in Weka. Many iterations of this process were undergone due to the process of feature engineering, in which we tried an array of techniques, including but not limited to taking the log of, squaring, normalizing the data, and building a variance-based classifier.

Results

Through a feature engineering process, we found that taking the log base 10 of the data generally improved the accuracy across the board, while building a variance-based classifier, cross-correlating the time series data, squaring the data, and more did not improve accuracy. Four machine learning algorithms - Simple Logistic, J48 Tree, Random Forest, and Best-First Tree - stood out, and the J48 Tree yielded the best results, with 70.83% accuracy in two out of the three subjects and 53% accuracy in the other whilst taking the log of the data.

Conclusions/Discussion

We were able to achieve satisfactory accuracy and meet our goal in two out of three subjects where data collection was optimal. Limitations of this experiment include the EEG Machine's simplicity, as well as the feature extraction methods used in pre-processing the data. We conclude that these results are promising and pave the way for future experiments, focused on improving accuracy and developing real-time data processing methods, to improve upon our work. Potential applications include helping people with Locked-In Syndrome (complete paralysis) communicate, and replacing keyboards with wearable headbands for communication in the future.

Summary Statement

The focus of this project was to determine what letter a subject is thinking about based on his/her brain waves, thus creating a thought-to-type interface.

Help Received

My school's research teacher acted as a mentor for the project. My father advised me on the algorithms portion of the experiment.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Prem M. Talwai

Project Number

S1425

Project Title

Model Order Reduction of Cell Signalling Pathways: An Investigation of the Invasive Mechanism of Ebola Virus

Abstract

Objectives/Goals

To formulate and implement a mathematical algorithm which reduces the dimensionality of kinetic models of cell signalling pathways while still preserving their most important dynamic features.

Methods/Materials

This algorithm (executed at the command-line interface through a Python interpreter) combines the techniques of proper orthogonal decomposition, trajectory piecewise-linearization, and Krylov subspace reduction in an effort to identify a best-fit subspace for the model trajectories that accurately approximates the mapping between the model input and outputs. The algorithm will be implemented on a kinetic model of the IFNG-JAK-STAT-EVP24 network, which will be constructed using the Simmune Modeler and Simulator. Once the reduction procedure has been successfully executed on the kinetic model, correlations between the reduced and full-order bases will be analyzed in order to reveal novel dynamical relationships between the species participating in the signalling cascade.

Results

The algorithm successfully reduced the dimensionality of a kinetic model of the IFNG-JAK-STAT-EVP24 network from 45 to 5. In addition, the algorithm revealed strong positive correlations within four groups of participating species, which divided the mechanism into four distinct temporal phases.

Conclusions/Discussion

The developed model reduction algorithm enables cell biologists to effectively employ kinetic models to analyze the mechanisms of complex diseases. Further work may involve obtaining an a priori error estimation of the accuracy of the reduction procedure and incorporating the algorithm as an accessible front-end tool in the Simmune graphical user interface.

Summary Statement

This project designs a mathematical algorithm for reducing the dimensionality of kinetic models of cell signalling pathways.

Help Received

Participant in National Institutes of Health Summer Internship Program. I was provided access to the Simmune software suite by the Computational Biology team; Drs. Martin Meier-Schellersheim and Bastian Angermann provided invaluable support and guidance.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s) **Raam M. Tambe**	**Project Number** **S1426**

Project Title

The Effect of Neutral Agents on the Expansion of Ideology in a Hypothetical Social Space

Abstract

Objectives/Goals

The majority of research in the field of evolutionary game theory uses computer models to predict the effects of ideas, values, and beliefs on a given population. This experiment attempts to incorporate the frequently documented "bystander effect" in existing computer models by creating a "neutral agent" that does not adopt any values or beliefs. The goal of this project is to ultimately refine the models that are currently used to collect such data by making them a more accurate representation of the real world.

Methods/Materials

Using Netlogo, a 50 by 50 grid is created to model a hypothetical social space, wherein each cell is an agent that that can choose among a neutral, passive, or aggressive strategy. Each cell is awarded points based on a set of predetermined values and their neighbor's strategies, then adopts the strategy of the highest performing neighbor. An "Evolutionary Stable Strategy" is determined when it is no longer possible for cells, whether passive aggressive or neutral, can no longer outnumber the cells of the currently dominant trait. This model is contrasted with a more common model that exclusively allows for aggressive and passive strategies. Data was collected across 3 tests with 30 trials each.

Results

Across three tests, with thirty trials each, the average distance in from the average rate of convergence, or the amount of cycles in the program, is 1.40 ticks for the model including neutral agents, compared to an average distance of 2.86 ticks for the model not including neutral agents. The model including neutral agents had a distance from the average expected cell count of 22.48, resulting in an average difference of 0.002 percent.

Conclusions/Discussion

These results indicate that modeling a social space to include neutral agents is feasible and can increase the statistical significance of data by reducing the variance per individual trial. This project can help inform other research concerning the expansion of ideas, whether it be focused on political polls, or likelihood of genocide.

Summary Statement

This project designs a new method of analyzing the ways in which ideas, beliefs, and values interact with individuals by trying to replicate a "bystander effect", and tested the model by comparing its accuracy to the current one.

Help Received

I received help from my father who works at USC.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Utkarsh Tandon

Project Number

S1427

Project Title

Signal-Processing Driven Machine Learning Algorithms for Parkinson's Progression and Tremor-Based Disease Diagnostics

Abstract

Objectives/Goals

This project develops a system for rapid hand tremor analysis to diagnose neurological disorders and conduct stage classification of Parkinson's disease (PD). Lacking proper quantification, neurological diseases are often insufficiently supervised, contributing to a death rate of 6.8 million people annually. Hence, this study focuses on quantifying neurological disorders through signal-processing based algorithms (FFT, PSD) that conduct analysis of associated hand tremors. A trained Random Forests machine-learning model (QWK accuracy goal>0.85) allows for predictions of PD severity through passively collected tremor data from the engineered wearable device, therefore creating an autonomous diagnostic tool for multiple neurological disorders.

Methods/Materials

Methodology of this study is split into three segments: Active analysis, Passive analysis, and Integration of the Python algorithms in an iOS based diagnostic system. Through Fast Fourier Transforms (time domain to frequency domain conversion via Euler derivation) and Peak Detection Algorithms, features of dominant frequency and average amplitude are extracted from 10-second hand tremor examination. These features populate a binary classifier capable of distinguishing between multiple neurological disorders (Essential Tremor, PD, Multiple Sclerosis, etc.). The passive analysis system utilized Power Spectral Density estimates to construct passive tremor trends (used to train RF module for PD severity prediction) in Michael J. Fox foundation sponsored training data.

Results

The Random Forest ensemble was tested through a Quadratic Weighted Kappa k-fold cross validation, which evaluated inter-rater agreement (human gold standard vs. ML model). For 127 iterations the RF module outputted an average 0.864 kappa score, nearing perfect prediction accuracy of PD progression. The completed device was additionally validated through the use of real PD patients where 4 out of 6 were successfully diagnosed with Parkinson's disease solely based of the tremor trends detected on their hands.

Conclusions/Discussion

The results represent high accuracies for PD stage identification and validation for neurological disorder demarcation based on hand tremor. This mathematical analysis of tremor creates a novel method of disease diagnostic and quantification, especially useful in rural regions lacking trained neurologists to monitor involuntary-movement based diseases.

Summary Statement

This project creates a Random Forest based machine learning model and signal-processing algorithms to analyze hand tremor trends in order to diagnose neurological disorders and conduct Parkinson's disease stage classification.

Help Received

Dr. Brigitt Schuele from the Parkinson's Institute and Clinical Center provided information about PD progression and introduced me to the patients at the clinic; Dave Matsumoto from MakeXYZ provided assistance during the 3D printing process.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Jared A. Tramontano

Project Number

S1428

Project Title

Fuzzy Structures with Application to Differential Topology, Manifold Learning, and Specialized Concepts

Abstract

Objectives/Goals

The main purpose of the project is to introduce a notion of Differential Topology in the setting of Fuzzy Set Theory, introduced by Zadeh in the 1960's. As Fuzzy Set Theory is just a generalization of set theory, Fuzzy Differential Topology is just a generalization in the normal Euclidean setting. All concepts were developed with application to Manifold Learning in mind. Namely, one may consider fuzzy data manifolds in computer science.

Methods/Materials

The only materials needed for this project were a blackboard and chalk, various textbooks (notably Differential Topology by Hirsch), and academic papers.

Results

The standard setting of Differential Topology was extended to the notion of Fuzzy Topological Vector Spaces. Although slight attempts had been made previously, such attempts lacked rigor and depth. The new notions developed include higher dimensional fuzzy sets, Fuzzy Topological Separation Axioms, Fuzzy Differentiation, Fuzzy Atlases, Fuzzy Tangent Bundles, Fuzzy Cotangent Bundles, and Fuzzy Lebesgue Integration.

These notions give rise to a proper foundation for Fuzzy Differential Topology, where most concepts, such as homology, should have suitable analogies in these spaces.

Conclusions/Discussion

In terms of computer science, these notions have been implemented in algorithms in manifold learning that seem to better handle real-world data sets, using the notion of fuzzy set theory. Two students at MIT have handled the writing of said algorithm, and the three person paper detailing my notions of Fuzzy Differential Topology and their algorithms, will be submitted for publication this Fall.

In summary, proper generalizations of Differential Topology have been established, thus setting the foundations for future work in the area.

Summary Statement

Combining concepts from Differential Topology and Fuzzy Set Theory.

Help Received

Discussions with various professors at UCR and UCI, as well as students of MIT and Harvey Mudd.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s) **Devansh Vaid; Katia Williams**	**Project Number** **S1429**

Project Title

Simple Calculus: The Transformation of an Oxymoron into a Reality

Abstract

Objectives/Goals

Our objective was to invent a Calculus notation that is directly influenced by the context of common problems and graphical representations.

Methods/Materials

The materials we used included textbooks, articles, encyclopedias, pencils, and papers.

Results

A successful notation was created that received polarizing reviews from professionals and amateurs alike.

Conclusions/Discussion

This project will definitely be expanded upon and improved, but for a first step, it was a success.

Summary Statement

Our project is about making Calculus simpler and more applicable and understandable for the layman.

Help Received

Out mathematics teacher taught us the Calculus concepts, with their implementations, that we needed to be successful for this project.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Timothy J. Varghese

Project Number

S1430

Project Title

From Sums over Natural Numbers to Sums over Primes

Abstract

Objectives/Goals

Any sum over the naturals such as 1 + 2 + 3 . . . n = F(n) can be converted to a sum over the primes such as 2 + 3 + 5 + . . p_n = P(n) where p_n is the n-th prime. I started by proposing an Expand-Sum-Prune (ESP) heuristic in which P(n) is approximated as F (n ln n)/ ln n. ESP provides correct asymptotic results for sums of prime powers, duplicating a result of Salat-Znam. The goals of this project are:

1. To examine the hypothesis that ESP fails when any one term is too large a fraction of the whole sum.
2. To find new sums over primes never published earlier
3. When ESP fails, to find better summation methods.

Methods/Materials

1. Series: I studied several sums over primes including the alternating series (2 - 3 +5 -7 + ...), reciprocal sums (1/(2*3*5)+ 1 (3*5*7) +. . .), and sums of prime powers (2^2 + 3^2 + 5 ^2 + . . .)
2. All estimates were checked for accuracy using a Visual C program that uses the sieve of Eratosthenes to produce (and sum) all primes up to 1000000.

Results

1. Alternating sum: Consider A = 2 # 3 + 5. I provide a new estimate of |A| = 0.5 p_n with errors less than 2% for 500<n<78,401 by summing half the prime gaps using a modified ESP method. When my estimate was posed on Math Overflow (viewed 386 times, 11 votes +1 badge for "good question"), mathematicians felt my new result was "almost certainly true". However, using current sieve techniques they can only prove unconditionally that |A|< p_n /64. My method generalizes to alternating series of prime powers. I published a new series for alternating primes squared in the Online Encyclopedia on Integer Sequences (OEIS) as A240860.
2. Reciprocal sums: I prove that S=1/(2*3*5) + 1/ (3 *5*7) + ... converges and 0.0474 < S < 0.0475, published in the OEIS as A242187 using a Bound-Reduce that applies to the infinite series for e.
3. Better estimates for sums of prime powers: I found a better approximation than the Salat-Znam estimate using a balancing constant c. I found experimentally that the best values of c are roughly 0.6 for prime sums, 0.7 for squared sums, and 0.9 for cubed sums
4. New estimates from old: I found a new asymptotic estimate for prime products two at a time with 2% error, added to the OEIS as A024447.

Conclusions/Discussion

The hypothesis that ESP method fails if any term dominates (limit of ratio of largest term to sum does not tend to zero) is supported by results.

Summary Statement

As in Alladin where the peddler promises new lamps for old, I seek new series over primes from old series over integers, and new formulas derived from formulas for integers

Help Received

Neil Sloane,head of OEIS helped refine hypothesis, Erich Bach (Wisconsin) helped make program efficient, Father helped with program. Robert Oliver (Stanford) gave valuable suggestions.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Francisca Vasconcelos

Project Number

S1431

Project Title

Object Recognition Based UAV Control

Abstract

Objectives/Goals

The goal of the project is to have an AR Parrot Drone follow a person using computer vision techniques to recognize an object that he or she is wearing or holding.

Methods/Materials

Objects are detected using a UAV mounted camera. The distance between the object center of mass, in the image plane, and the image center is measured. This distance is used as an error signal to control the UAV velocity. The feedback loop is implemented with a proportional-integral-derivative (PID) controller. The computer vision techniques used for object detection are inspired by the attention mechanisms of human vision, exploiting a combination of bottom-up and top-down saliency cues to speed up recognition. The saliency mechanisms are implemented with a combination of the Harris interest point detector and an object detector cascade. They enable recognition at video frame rates.

Results

Experimental evaluation has shown that the goal of the project is feasible. The drone is currently able to recognize, follow, and stay a specified distance away from a patterned patch. It was found that, for the drone to achieve real time performance, a trade-off must be implemented between recognition accuracy and speed.

Conclusions/Discussion

Although computer vision has previously been used in UAV literature, the emphasis has been on motion computations (optical flow) for navigation and obstacle avoidance. The techniques now introduced enable the UAV to recognize specific objects and react to them. In particular, the UAV can track a person wearing an object, such as a piece of clothing or patch. Unlike current GPS-based methods, the behavior of the UAV can vary according to the object being tracked. UAV programming thus becomes as simple as changing your clothes (or pattern). The UAV is also much more precise in its interaction with the user and works both in and outdoors. Many applications could follow from this technology, including a UAV that behaves as an older brother that watches a child from above, a personal cameraman that follows an athlete on the field, an additional pair of eyes that allow a bike rider to "look around the corner," or a sitter that watches a pet as it roams around the park.

Summary Statement

The project is about the use of computer vision techniques to enable a UAV to follow a person and react to what he/she is wearing/holding.

Help Received

I used open source code for the low-level infrastructure of communicating with the drone and OpenCV for some of the computer vision functionality.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Conner J. Vercellino

Project Number

S1432

Project Title

Jabberwocky: Use of Unsupervised Stacked Autoencoders for Next Generation Spell Checking Software for Dyslexics

Abstract

Objectives/Goals

Jabberwocky is a next generation spelling and grammar checking that uses stacked deep unsupervised autoencoders to make corrections from actual understanding of sentence context. This creates a far more accurate spell checker, comparable to a human editor. Using unsupervised text, it is able use a huge amount of data, removing the need for structured training data that most spell checkers need. It is able to learn individual or group spelling patterns to further augment the corrections, via structured datasets specific to the targeted user group.

Methods/Materials

The program needs to use entire sentences in order to be able to use the context of the sentence to help decipher what word the user intended to use. I looked into unsupervised models for neural networks. I decided on an noising filtering autoencoder based system. An noising filtering autoencoder is a type of neural network, whereby it tries to reproduce the input it is given without noise. Autoencoders have never been used for text correction making this project's approach to the problem of grammar and spelling correction very novel.

Results

Currently the network can make extremely accurate corrections on simple sentences. This proves that both an unsupervised neural network can work for spell correction, and that noise filtering does work on natural language. The program has been able to fix improper grammar and correct spelling on repeatedly mutated sentences. However, the network has yet to prove that it can generalize on new complex sentences that it hasn#t seen before.

Conclusions/Discussion

Jabberwocky has many uses, one such being that sites like Yelp, Airbnb and Wikipedia could be completely free of spelling and grammar errors, leading to the site#s content being more readable and professional, without effort of the site#s users. Search sites could use Jabberwocky on their queries, thereby removing the need to make their searches need to watch for improper use of spelling and grammar. Lastly, Jabberwocky, while currently trained for correcting text, could theoretically be adapted to facilitate code correction. This would be an extremely useful, as it could learn from existing code and it would be a huge boon to programmers.

Summary Statement

Using unsupervised stacked noise filtering autoencoders for creation of a extremely powerful NL correction algorithm (primarily aimed at Dyslexics).

Help Received

Classmate helped with the dataset loader; Andrew Saxe helped explaining about neural network training and math; Jen Selby helped with time management.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Yinghao Wang

Project Number

S1433

Project Title

A HOG-AdaBoost Method for Stop Sign Detection

Abstract

Objectives/Goals

Creating a new method of stop sign detection that generates more accurate results than a heavily used traditional method (neural network)

Determine a equation relating the size of stop sign with the distance between the stop sign and the camera

Methods/Materials

Collect one thousand positive and one million negative samples of stop signs. Histogram of Oriented Gradients (HOG) features were extracted from the training set and the Adaboost algorithm was applied to train the strong classifier with different number of weak classifiers.

Test the HOG-Adaboost and Neural Network algorithms. Compare the results of the neural network algorithm to the results of the Adaboost algorithm.

Results

The accuracy of my HOG-AdaBoost algorithm was 99.832% while the accuracy of the neural network algorithm was 99.74% with the same feature set, implying that my HOG-AdaBoost algorithm had a better detection accuracy. My HOG-AdaBoost algorithm also ran much faster than the neural network algorithm. I also found an equation relating the size of the stop sign with the distance between the stop sign and the camera. The equation had an average error of less than 0.13 feet compared with the ground truth.

Conclusions/Discussion

In my hypothesis, I stated that my HOG-AdaBoost method would detect stop signs more accurately and faster than the neural network algorithm. Also, I stated that if I am given the size of a stop sign, I will be able to find how far away the stop sign is.

According to my data, my hypothesis was supported. My HOG-AdaBoost algorithm had a better detection accuracy than the neural network algorithm due to the effective strong classifiers. With the same training set of 324 features, my HOG-AdaBoost algorithm also ran 30 times faster than the neural network algorithm because the neural network had much more features and weights to deal with.

Summary Statement

Creating a new method of stop sign detection that is more accurate and efficient than the traditional neural network method

Help Received

Dr. Wang helped guide the project

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Kevin Wu

Project Number

S1434

Project Title

A Novel Framework and Algorithm for Ultra-fast Protein Shape Comparison

Abstract

Objectives/Goals

Scientists study the structural similarity of proteins for many purposes, such as developing new medicines, finding protein evolutionary relationships in the protein families, etc. With more than 100,000 protein structures already discovered, the comparison of protein structures becomes more complex and time consuming. Current protein comparison takes about 12 hours to compare one protein structure against all protein structures.

A fast comparison algorithm to filter out any proteins that are definitely similar or not similar to a given protein is an effective approach to speed up the process.

My engineering goals/objectives are:

1. Design and develop a framework to allows scientists to develop and test new protein comparison algorithms more conveniently and efficiently
2. Design and develop new algorithms to achieve fast and accurate protein shape comparisons.

Methods/Materials

Step sequence of Framework:

The Framework procedure is as below:

1. Pull the Query and Rep protein data stored locally of over the network.
2. Extract the 3D points of the Query and Rep proteins.
3. Apply the Fingerprinting method to all of the proteins.
4. Applies the Scoring method, comparing Query proteins with Rep proteins.
5. Store the pairwise scores.
6. Use the pre-saved TM scores to plot the statistics results based on the scores.

Algorithm Method

Input: 1. Query Protein List; 2. Selected cluster of Representative Proteins.

Output: 1. Pairwise Scores; 2. Statistics Plot.

Materials: 1. Super Computer Center of UCSD. 2. Access to the RCSB PDB. 3. 27,000 Representative Proteins Data. 4. Facility for conducting the research: Skaggs School of Pharmaceuticals. 5. A laptop computer. 6. Java Programming Language. 7. Eclipse IDE for Java.

Results

1.The framework is able to run multiple algorithms at the same time. Each algorithm yields one plotted result set. This provides the developers a more flexible and efficient way so that one time execution can output multiple results for multiple algorithms.

Summary Statement

The purpose of fast protein shape comparison is to speed up the development of new medicines, predict new protein functions, and discover evolutionary relationships of proteins.

Help Received

I would like to thank the Skaggs School of Pharmaceuticals at University California of San Diego for letting me to use Protein Data Bank to conduct my research under the supervision of Dr. Peter Rose.

CALIFORNIA STATE SCIENCE FAIR
2015 PROJECT SUMMARY

Name(s)

Sonia Sachar

Project Number

S1499

Project Title

A Systems Biology Approach to Optimize Prediction of Efficacy of Pathway Targeted Therapies in Cancer

Abstract

Objectives/Goals

One of the most challenging problems in the field of medical science is predicting the drugs that will work for specific individuals diagnosed with varying cancer types. Today doctors have many anti-cancer therapeutics to choose from; however, doctors are only able to predict which drug would work primarily based on toxicity levels. The objective of this project is to develop a software tool to effectively test my hypothesis that drugs targeted to a specific pathway are only effective if that pathway is up regulated. These up-regulated pathways can be identified using a data driven approach by analyzing the interaction between proteomic, transcriptomic, and drug response data.

Methods/Materials

I focused on Non-Small Cell Lung Cancer for this research project and used corresponding lung cell data. My procedure consists of three parts: data discovery and analysis, statistical inference, and dynamic visualization software. In data discovery and analysis, relevant data was collected from multiple sources and integrated into a data cube. For statistical inference, various methods such as paired p-values, probability density bell curves, and scatter plots were used to compare fifty-eight common pathways across a pair of cell lines where the same pathway-targeted therapy worked on one cell line and not on the other. Finally, dynamic visualization software was developed to view dynamic models of these pathways.

Results

Of the common pathways, 94.8% were unregulated in the same cell line where the drug had a positive response. Thus I validated my hypothesis and developed an approach to view various biological data such as transcriptomic, proteomic, and drug data as a system and effectively determine the correlation between these data sets as well as visualize dynamic models with the integration of bimolecular signaling networks, protein-protein interactions, and transcription factor regulation.

Conclusions/Discussion

I effectively developed a user-friendly data analysis and visualization software that creates a full pathway analysis for a specific patient based on the patient#s proteomic and transcriptomic tumor analysis. This software can successfully help clinicians narrow down the number of drugs that have a higher probability of working for an individual cancer patient.

Summary Statement

Using translational systems knowledge to dynamically map bio-molecular pathways and data analytics, I developed a data analysis and visualization software that effectively creates a full pathway analysis for a specific patient based on the

Help Received

Dr. Parag Mallick at Stanford helped me understand the various biological datasets, reviewed my methodology and verified the viability of this approach.