Report on
Georgia Tech Math and Computer Science Symposium

Conference Synopsis

The Mathematics and computer science symposium's purpose was to introduce Graduate students in non-analytical fields to students in a low-stress environment. The conference was held over one-day 9am-5pm at Georgia Tech. The conference consisted of four content sessions: Introduction to Python, Introduction to Linear Algebra, Introduction to Statistics, and finally introduction to Machine Learning. In addition to the content section, the conference had two interactive breakfast and lunch sessions used to facilitate a stress-free environment and concluded with a poster session showing machine learning applications.

Conference Roles

Conference coordinator: Christin Salley
Content Instructor: Breanna Shi
Teaching Assistants:

Intro to Python	Suhani Madarapu	Arushi Aggarwal	Harini Mudradi	
Intro to Linear Algebra	Avinash Palliyil	Daniel Lee	Arushi Aggarwal	
Intro to Statistics	Anand Tsogtjargal	Avinash Palliyil	Arushi Aggarwal	Noah joseph
Intro to Machine Learning	Edward Chong	Oluwaferanmi Akande	Arushi Aggarwal	Asmita Lagwankar

Content Development Lead: Arushi Aggarwal
Content Development Team:Suhani Madarapu, Oluwaferanmi Akande, Edward Chong, Daniel Lee, Avinash Palliyil, Tejaswini Ramkumar Babu

Instructional Strategy

The purpose of the conference was to facilitate better understanding and relationships between analytical and non-analytical collaborators. As interdisciplinary research becomes more common, it is necessary for both sides to understand the value and contributions of each other's work. It is very common that colleagues that work more commonly in a wet lab or otherwise

non-analytical environment are apprehensive about learning analytical topics. It is becoming more common that every researcher can perform and understand basic analytics on their data. The main goal of this conference was to have students with no analytical experience leave with a basic understanding of Machine Learning.
Our approach was two-fold: minimize lectures essential content and provide working time for participants to practice. Each section was one hour long and split in the following way: 20 minutes of lecture, 30 minutes of practice in Colab, 10 minute break. Every lecture was intentionally ten slides long with one image describing a fundamental topic. The ten subjects per lesson were chosen for the relevance to the overarching goal of preparing a non-technical student to understand basic machine learning code. This means the topics in the courses introduction to python, introduction to linear algebra, introduction to statistics were not chosen to give a broad overview of those topics. Instead the topics were chosen if they were necessary to understanding a basic machine learning script. The predictability of the course structure was to aid in creation of a low-stress environment.
The content for the practice sections was developed by computer science undergraduates and masters as a research project under the advisory of Breanna Shi. Students who generated the course materials also assisted as lesson teaching assistants to assist in participants understanding during the 30 min practice section. Additionally, the teaching assistants were available during the lecture in a remote section of the classroom, so that confused students might feel comfortable to discreetly ask a question at any point during the conference. This serves the participants because although the instructor is always available for questions during the conference, it can be stressful to ask questions to a person you perceive as more senior. In addition, the masters and undergraduate teaching assistants are closer to their ‘learning stage’ in terms of machine learning content. Thus, their perspective is likely extremely relevant to a new learner such as our participants.

Instructional Tools

During the conference, we utilized slack and Colab notebook as educational tools for our lesson. The slack was used for answering questions and connecting with other participants. The Colab notebook was used to facilitate our practice session. This was an extremely vital tool as using an online platform for coding illuminates hardware compatibility issues. Participants being informed of the educational strategies used in this conference for transparency. This serves to set expectations for the course and allows the participant to feel confident while learning difficult topics. Surveys were collected to assess the effectiveness of our approach (results will be outlined later in this document).

Poster Abstracts

After the conference lessons, we held a small poster session showcasing an active Machine Learning project on-going at Georgia Tech. The purpose of this showcase was to inspire the participants with what new research they could accomplish with their new found Machine Learning Knowledge.

Here is a list of our presentations:

Title:

Exploring Sampling Methods for Addressing Imbalanced Class Distribution in Credit Card Fraud Detection

Speaker:

Arnav Hiray

Abstract:

In the context of credit card fraud detection, imbalanced class distribution is a common issue that can negatively impact the performance of machine learning models. In this study, we compared various sampling methods to handle imbalanced classes and evaluated their effectiveness in improving model performance. The methods evaluated included random undersampling, random oversampling, SMOTE (Synthetic Minority Over-sampling Technique), ADASYN (Adaptive Synthetic Sampling), and a combination of random undersampling and SMOTE. Our experimental results showed that while the combination of random undersampling and SMOTE tended to outperform the other methods in terms of improving model performance, the differences were not always statistically significant. Additionally, we found that the best-performing sampling method varied depending on the specific machine learning algorithm used. Overall, our study provides insights into the effectiveness of different sampling methods for addressing imbalanced class distribution in credit card fraud detection, but further research is needed to better understand the nuances and trade-offs of these approaches.

Title:

Compiler-RL: Reinforcement Learning Approaches to Phase Ordering

Speaker:

Adithya Vasudev

Abstract:

There is already research in applying RL techniques to phase ordering, targeting a wide variety of resulting bene- fits, such as smaller memory foot- print, lower instruction count (and consequently, smaller executable file size), as well as effective speed of executable. The problem can be formalized as learning a function that maps an Intermediate Representation (IR) of a program to a weight for each possible optimization pass and optimizing this function based on a reward. We wish to investigate further using Deep RL to improve upon existing results. In this paper, we provide an overview of tackling a problem using Deep RL specific to the ProGraML [4] observation space, which represents code in graph form. To supplement this graphical approach to code representa- tion, we further propose MultiProGraML, an augmenta- tion of ProGraML that encodes multiplicity information to better encapsulate looping and repetition in code. Finally, we present two approaches to tackling this phase or- dering problem involving reinforcement learning: graph- based Deep Q-Networks (DQN) and Proximal Policy Or- dering [12]. We show that DQN models, selecting one pass at a time from a policy, do perform better than both the base LLVM compiler and randomly selecting passes, and that PPO models are excellent when it comes to selecting passes, performing significantly better than LLVM’s preset order.

Title:

AutoDifferentiation-Accelerated Methods for Finding Relative Periodic Orbits (RPOs) in Turbulent Flows

Speaker:

Stanislaw Kowalski

Abstract:

The problem of finding exactly recurrent solutions to nonlinear partial differential equations is a central problem in the field of turbulent flows, particularly controlling and understanding the statistics of their behavior. We design a framework using a Google research machine-learning backend, JAX, that provides the capability to propagate full precision derivatives through many thousands of timesteps in any fluid system solver with constant overhead. This allows advanced matrix-free gradient descent and adjoint looping algorithms, such as nonlinear Newton-Krylov subspace iteration, to be used to find fluid solutions with arbitrary properties, such as RPOs in Kuramotov-Sivashinsky laminar flame fronts or magnetically-driven Kolmogorov Flow, which we demonstrate.

Title:

Pose estimation and Trajectory analysis of fish videos with deep learning

Speaker: Asmita Lagwankar

Abstract:

Understanding social behavior can be difficult, but computational methods raise an interesting possibility to automate a process which has traditionally required human-observation. Cichlids are a model species for studying complex social behavior such as mating, schooling, feeding, and aggression. In our study, we will predict behavior on three 1-hour videos taken of the cichlid species Mchenga conophoros (MC) and yellow head (YH). We will use popular animal behavior software DeepLabCut (DLC) to learn the pose-estimation of our fish in a given frame and combine these observations into tracks which can be further analyzed into stereotyped and non-stereotyped behaviors. Since behavior movement patterns will have sequence variation across time-series positional coordinates, we will need to account for these variations in the generation of class patterns prior to applying an unsupervised classifier. Behavior patterns will then be classified using unsupervised classification algorithms such as PCA, UMAP, TSNE.

Title:

VIP FishStalkers: Using Machine Learning Pipeline to Quantify Behavior

Speaker:

Breanna Shi

Abstract:

Complex social behaviors are essential to survival and reproduction, but the existing methods are ill-suited to give the researcher agency to conduct rigorous and controlled experiments. Inducing social behavioral responses is difficult, relying on interactions with multiple animals that behave in uncontrolled and sometimes unreliable ways. Additionally, stimuli from animals are multi-modal including multiple visual cues that are difficult for the experimenter to control.

Teasing apart the multivariate conditions of an animal behavior requires a combinatorial number of experiments that increases exponentially with respect to the complexity of the behavior and its stimuli. I will use object detection and tracking algorithms to accurately and reproducibly quantify mating and feeding behavior of lake Malawi cichlids.

Advertisement

We put up flyers around campus, specifically in the schools of sciences, and common study areas such as the grad student lounge, the clough undergraduate learning commons, and around the student center. We sent emails on mailing lists, and advertised through discord servers, groupme, and grad student group chats.
Note, we advertised as an opportunity especially for diverse participants, but did not place any demographic restrictions on attendance. We made significant efforts to advertise to organizations and institutes on campus that support diverse student populations.

Budget

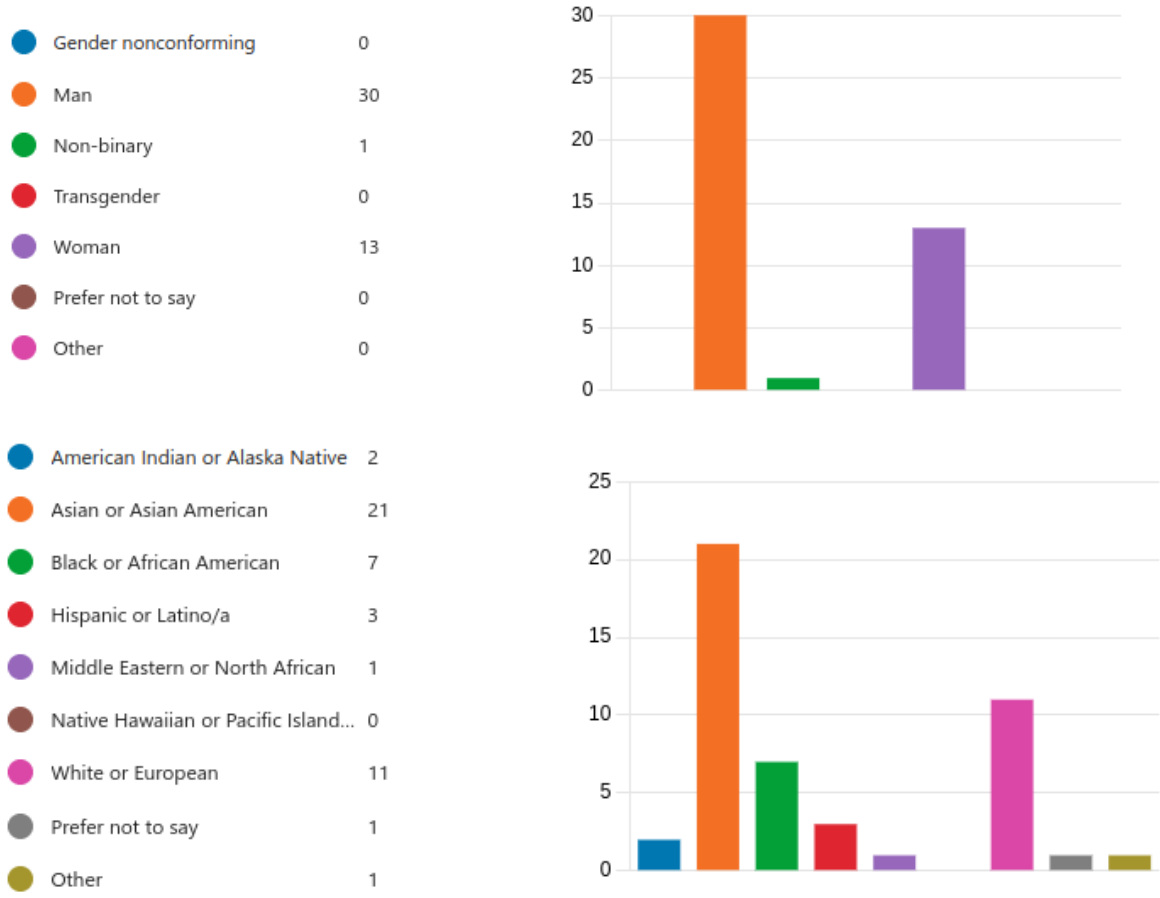
Item	expense
Georgia Tech Catering	$1275
Room Rental	$180.60
Speaker honorarium	$544.40
Conference Organizers Stipend	$1000 ($500 each)
Total	$3000

Participants

Pre-Conference

We had 45 respondents register for the conference who self-identified at the PhD (17), Masters (10), and undergraduate (18) level.

The following shows how participants self-identified demographically:



We asked the registrants to (optionally) share why the symposium is impactful for them. Here is some of the responses we received:

- “I would like to enter the bioinformatics field and think this conference would be helpful! ”
- “These skills are applicable to the third chapter of my PhD.”
- “I’m interested in applying machine learning and AI to cell therapy by creating high precision medicines ; however, I lack experience in ML and computer science and am hoping to learn more about them through this symposium.”
- “I am very excited about this because I have recently been thinking about changing my career to focusing on this field and would love to learn about what is involved in it.”
- “I would like to grow in my computational skills, specifically in the field of machine learning. I have always had a small fear for the phrase “machine learning,” largely because I did not know what it was. This semester, I am taking a class that teaches some theory behind machine learning applications in biosciences and I have grown to appreciate the large-scale impact ML can have on my field. I would love to attend this symposium and learn more about ML in a low-stress environment, while connecting with fellow women who are in the same space.”
- “Interested in STEM, but nervous about not fully understanding.”
- “I'm interested in seeing how others are applying machine learning in nonstandard ways, presenting my project, and seeing how I could apply ML to my fields!”
- “I applied to a PhD program”

During Conference

One challenge for assessment is that we always had lower survey response numbers than actual participants due participants coming and going with whatever topics were relevant to them. We are calculating the number of participants by the number of non-facilitating members who joined the slack group. We estimated there were 15-20 participants at a given time during the conference. This was only given to participants. Regardless, we can still draw some useful insights.

Pre-Conference Baseline Survey (after Breakfast)

We received 13 responses. Average score per statement: (With 1 meaning the statement is untrue and 10 meaning the statement is completely True for the participant)

1. I am feeling of stress or anxiety : 4.69
2. I am confidence in my ability to understand the ideas taught: 6.77
3. I am concerned that the content will be too advanced for my understanding: 5.69

Introduction to Python Feedback Survey

We received 9 responses. Average score per statement: (With 1 meaning the statement is untrue and 10 meaning the statement is completely True for the participant)

1. I had feelings of stress or anxiety during the lecture: 3.44
2. I had feeling of stress or anxiety during practice session: 3.56
4. I was able to understand the ideas taught: 8.22
5. I was provided helpful assistance during the session: 9.11
6. The content was NOT too advanced for my understanding: 8.00

Introduction to Linear Algebra Feedback Survey

We received 4 responses. Average score per statement: (With 1 meaning the statement is untrue and 10 meaning the statement is completely True for the participant)

1. I had feelings of stress or anxiety during the lecture: 4
2. I had feeling of stress or anxiety during practice session: 4
3. I was able to understand the ideas taught:7.5
4. I was provided helpful assistance during the session: 8.75
5. The content was NOT too advanced for my understanding: 7.75

Introduction to Statistics Feedback Survey

We received 7 responses. Average score per statement: (With 1 meaning the statement is untrue and 10 meaning the statement is completely True for the participant)

1. I had feelings of stress or anxiety during the lecture: 4.29
2. I had feeling of stress or anxiety during practice session: 5
3. I was able to understand the ideas taught: 7.29
4. I was provided helpful assistance during the session: 8.43
5. The content was NOT too advanced for my understanding: 7.29

Introduction to Machine Learning Feedback Survey

We received 3 responses. Average score per statement: (With 1 meaning the statement is untrue and 10 meaning the statement is completely True for the participant)
I had feelings of stress or anxiety during the lecture: 3.67
I had feeling of stress or anxiety during practice session: 3
I was able to understand the ideas taught: 8.33
I was provided helpful assistance during the session: 7.67
The content was NOT too advanced for my understanding: 8

Follow-Up

Following the conference, we encourage participants to connect and share their struggles/ triumphs as they continue independently studying machine learning. In the summer, we are inviting the participants to an advanced and beginner weekly journal club to continue collaboration across experience levels on our campus. We hope to continue to improve our teaching approach and facilitate this conference on an annual basis.


COLLEGE OF SCIENCES &
COLLEGE OF ENGINEERING
Mathematics and Computer Science Symposium
CO-ORGANIZED BY: BREANNA SHI &
CHRISTIN SALLEY