Science Atlantic
MSCS | Mathematics & Statistics and Computer Science Conference

The 46th Matematics Statistics and Computer Science, Science Atlantic Conference

October 13-14, 2023

Charlottetown

University of Prince Edward Island

Atlantic Association for Research in the Mathematical Sciences

Contents

1 Welcoming Messages

Welcome to the 46th edition of the Science Atlantic - Mathematics, Statistics and Computer Science (SA MSCS) Conference held in Charlottetown, PEI.

This conference is a continuation of the Atlantic Provinces Council on the Sciences Mathematics, Statistics and Computer Science conference that was started more than 45 years ago in 1978.

In this booklet we include the program of the conference. This year have already registered more than 120 people attending the conference. The program includes 40 abstract submission (39 talks), both graduate and undergraduate students, a programming competition, a mathematics problem solving competition, and three invited speakers. We have included a copy of the program for this event in this booklet. A live version is also available on the conference website. It will be a very busy weekend, and hopefully an enjoyable one as well. Once again, welcome to UPEI. My very best wishes to each of you for a stimulating and enjoyable conference.

Cezar Câmpeanu

Chair,
Organizing Committee of the Conference

2 General Information

- For the duration of the conference, participants may park in any UPEI parking lot posted as Reserved Parking after 5 p.m. Friday. Please do not park in lots posted as Designated Parking. The back parking lot (Lot "B") will be closest to the conference venues. Before 5 p.m. Friday, you can park at the Farmer's Market or the Superstore. Parking on campus before Friday at 5 p.m. is at your own risk.

- On Saturday, all of the activities will be held in MCDH (building # 12 on your Campus map). Please enter from the Est side (the side nearest Belvedere Avenue).

- All keynote lectures and opening will be held in the MCDH 242. (Building # 12 on the Campus map).

- Participants in the CS programming competition should check-in at MCDH 237 (the hallway) by 11:45 p.m. The competition will take place in CASS (building # 10 on the Campus map) at the first level (Rooms # 102, # 104 and # 105).

- Participants in the Math Competition should be at the MCDH 237 (the hallway) by 1:00 p.m. The rooms allocated for the competition are # 328 and #329.

- The Pizza party for competition participants is McDougall Hall (MCDH)-Schurman Market Square - 248 at 6 p.m.

- The Reception on Friday evening, is also in McDougall Hall (MCDH) - Schurman Market Square - 248 at 6 p.m.

- lunch on Saturday will be on Wanda Wyatt Dining Hall - Dining Hall - 126 (Building # 18 on the Campus map).

- The Banquet on Saturday afternoon will be held in the WA Murphy Student Centre - McMillan Hall - 110 (Building # 4 on the Campus map).

3 Acknowledgements

We wish to acknowledge the support of the following groups and thank them for their support and participation in the 2023 Science Atlantic Mathematics, Stattistics and Computer Science conference.

- The Science Atlantic
- The Atlantic Association for Research in the Mathematical Sciences (AARMS)
- The School of Mathematical and Computational Sciences,
- The Dean of Science, Nola Etkin,
- The Vice President Academic, Greg Naterer,
- The Vice President of Graduate Studies, Marva-Nixon,
- Meetings and Conventions PEI.

We would also like to thank the students, faculty, and staff of the School of Mathematics and Computational Science, all volunteers, judges, UPEI - Computer Services, UPEI Conference Services, and DANA Hospitality for their assistance.

Conference Schedule

Friday, October 13, 2023

Time	Activity	Location
12:00-3:00	Registration	MCDH 237
12:00-12:30	CS Programming Competition briefing	MCDH 242
1:00-6:00	Computer Science Programming Competition	CASS 101, 104, 105
2:00-5:30	Mathematics Competition	MCDH Room 328
2:30-3:30	Joint meeting of Math/Stats and CS committees	MCDH 242
3:30-5:00	Math & Stats committee meeting	MCDH 243
3:30-5:00	CS committee meeting	MCDH 246
5:00-7:00	Registration	MCDH 245
6:00-7:00	Pizza Party	MCDH-Schurman Market Square
7:00-7:15	Official Welcome and opening remarks	MCDH 242
7:15-8:15	Blundon Lecture: Franklin Mendivil *Sets of sums and sums of sets*	MCDH 242
8:15-9:15	Reception	MCDH-Schurman Market Square

Saturday, October 14

Time	Activity	Location
8:00-11:00	Registration	MCDH 245
8:10-10:30	Undergraduate Talks in Mathematics/Statistics	MCDH 328 & 329
8:10-10:30	Undergraduate Talks in Computer Science	MCDH 246
10:30-11:00	Refreshment Break	MCDH-Schurman Market Square
11:00-12:00	Field Lecture: Joanna Mills Flemming *Portholes into an Underwater World: Data and Models Unveil Secrets of the Sea*	MCDH 242
12:00-1:30	Lunch Break	
1:30-3:30	Undergraduate/Graduate Talks in Mathematics/Statistics	MCDH 328 & 329
1:30-3:30	Undergraduate/Graduate Talks in Computer Science	MCDH 246
3:30-4:00	Refreshment Break	MCDH-Schurman Market Square
4:00-5:00	Sedgwick lecture: Kathleen Fraser *Ethical AI: What does it mean, how can we do it, and why should we care?*	MCDH 242
5:30-6:30	Conference Banquet	WA Murphy Student Centre-McMillan Hall-110

4 Detailed Program for Saturday, October 14

Talks are held in three parallel sections.

5 Session I[4]

MCDH 246

Time	Speaker	Title	
	Spandana Chereddy	An Alignment-Free Approach to DNA Sequence Classification and Microbial Adaptation Analysis	U-CS
8:10-8:30	Aakanksha Khandwaha	Analysis of racial and gender bias in MIMIC-IV clinical notes via natural language processing	U-CS
8:30-8:50	M. A. Agowun	A Two Stream GNN for Point Cloud Classification	U-CS
8:50-9:10	S. Wadhwa	Unveiling the Potential of Ubuntu Touch OS: Privacy-Centric Features, and Private Server Integration	U-CS
9:10-9:30	Baxter Madore	A Submodelling-Based Approach to Expected Points in North American Football using Scikit-Learn	U-CS
9:30-9:50	Xinyi Li	Saving energy for multitask soft real-time system using RL and DVFS	U-CS
9:50-10:10	Lai T Uyen	Verification of Heaps' law in large language model emulated text with implications for energy-efficient query processing	U-CS
10:10-10:30	Wenwen Wang	Fourier Transform-based Unsupervised Clustering of DNA Sequences	U-CS
1:30-1:50	Charanpreet Singh	Enhancing Potato Crop Disease Detection and Classification Using Deep Learning and Robotic Imaging	G-CS
1:50-2:10	Ethan Heavey	The Art of Contradiction Detection	G-CS
2:10-2:30	Riya Pandey	Exploring the Genomic Signatures in a lower Dimensional Space	G-CS
2:30-2:50	Mingyue Guo	A Solution to The Four-Dogs Pursuit Problem by a System of Ordinary Differential Equations and a Description of Instantaneous Position During the Process Based on the Solution	U-MS
2:50-3:10	Hudson X. Forance	The Proper Spherically Symmetric Frame in Teleparallel Geometries	U-MS
3:10-3:30	Timothy Power	An Extension of the Solow-Swan Growth Model	U-MS

---

[4] All Computer Science talks and two Mathematics and Statistics undergraduate talks.

6 Section II[5]

MCDH 328

Time	Speaker	Title	
8:10-8:30	Kevin Saville	The Science of Selection: Predicting NBA All-Stars Through Analytics	U-MS
8:30-8:50	Anders Cornect	An Improved Algorithm for Gromovs Approximating Tree	U-MS
8:50-9:10	Ethan Saunders	Proper Circular Arc Graphs and Circular Robinson Spaces	U-MS
9:10-9:30	Andrew Allen	Reducing the T-count of quantum C^n-NOT gates	G-MS
9:30-9:50	Anson Green	QFASApp: A Friendly Interface for QFASA	U-MS
9:50-10:10	Griffin Bartlett	Submission Ideas in Orlicz Spaces	U-MS
10:10-10:30	Heather MacTavish	Modelling the Effectiveness of Non-Pharmaceutical Interventions on Influenza A & B	U-MS
1:30-1:50	Erin Hughes	Domination Polynomials of Graph Operations	U-MS
1:50-2:10	Ethan Saunders	Misère Cricket Pitch	U-MS
2:10-2:30	Nimarjeet Bajwa	All oriented paths and cycles are unimodal	U-MS
2:30-2:50	Logan Pipes	Mixing Models for Domination Reconfiguration	U-MS
2:50-3:10	Samuel Sarria Hurtado	Maximum Nim with a Pass using Fractional Restricions	U-MS
3:10-3:30	Simon Maltby	Investigating the Spatial Distribution of the Underlying Risk Factors for Covid-19 in Canada	U-MS

---

[5] The Mathematics and Statistics Undergraduate Student Session

7 Section III[6]

MCDH 329

Time	Speaker	Title	
8:10-8:30	Aadesh Nunkoo	Well-being of Blind or Low Vision Canadian Adolescents	G-MS
8:30-8:50	Daniel Teixeira	Dualizability through 2-categories	G-MS
8:50-9:10	Saurav Neupane	Causal Effect of Factors Impacting Participation in Music Education in Canadian Adolescents	G-MS
9:10-9:30	Janet Cheng	Prediction models of mortality for people with cystic fibrosis	G-MS
9:30-9:50	Éloïse Soucy	Identifying Gambling Personae Through Machine Learning Algorithms	G-MS
9:50-10:10	Louis Bu	A Comparative Study on Optimization Methods for Maritime Transport Routing	G-MS
10:10-10:30	Joy Liu	Spatiotemporal Models for Exploring Variability in Scallop Condition across the Bay of Fundy	G-MS
1:30-1:50	Linh Dinh	Quaternionic methods in exact synthesis	G-MS
1:50-2:10	Jingyu Li	Exploring the Impact of Bubble Strategies on the Spread of Infectious Diseases	G-MS
2:10-2:30	Dylan Pearson	Modelling Virus Containment with Self-disseminating Vaccines on n-Dimensional Grids	G-MS
2:30-2:50	Claire Cui	Copula Mixture Regression Models for Multivariate Response Data	G-MS
2:50-3:10	Scott Wesley	Visualizing Qudit Controls With Sheets	G-MS
3:10-3:30	Ethan O'Connell	Validation of Earth System Models	G-MS

1. MCDH = MacDougall and Marion Hall
2. G-CS = Graduate Student Computer Science
3. G-MS = Graduate Student Mathematics and Statistics
4. U-CS = Undergraduate Student Computer Science
5. U-MS = Undergraduate Student Mathematics and Statistics

---

[6] Mathematics and Statistics Graduate Student talks

8 Keynote Addresses

8.1 The Blundon Lecture (Mathematics Keynote address)

Franklin Mendivil *Acadia University*

SETS OF SUMS AND SUMS OF SETS

Friday, October 13 7:15-8:15 pm in MCDH 242

Franklin Mendivil is a Professor with the Department of Mathematics and Statistic at Acadia University. Dr. Mendivil has a B.S. in Civil Engineering, and a Master and Ph.D. in Mathematics from Georgia Institute of Technology. His research area involves fractals and their applications in imaging. Dr. Mendivil has started working in topology, and wrote two papers on compactifications. He also has research done in genetic algorithms and other types of stochastic methods for global optimization. His research is reflected in an impressive number of published articles, as well as two books published in Springer.

8.2 The Field Lecture (Statistics Keynote address)

Joanna Mills Flemming *Dalhousie University*

PORTHOLES INTO AN UNDERWATER WORLD: DATA AND MODELS UNVEIL SECRETS OF THE SEA

Saturday, October 14 11:00 am -12:00pm in MCDH 242

Joanna Mills Flemming is a Professor with the Department of Mathematics and Statistic at Dalhousie University. Dr. Flemming has a B.Sc. from University of Guelph, a Master of Science from Technical University of Nova Scotia, and Ph.D. from Dalhousie University. Her scholarly interest involves the development of statistical methodology for data exhibiting spatial and/or temporal dependencies with a particular interest in what is important for marine ecology, and more broadly, environmental science. Example of her research topics are the development of statistical methodology for statistical tools for modelling marine animal movement data, and statistical methodologies and computational techniques for the modelling of complex ecological data, with several papers published on the results.

8.3 The Sedgewick Lecture(Computer Science)

Kathleen Fraser *Ottawa/Ontario, Canada*

ETHICAL AI: WHAT DOES IT MEAN, HOW CAN WE DO IT, AND WHY SHOULD WE CARE?

Saturday, October 14. 4:00-5:00 in MCDH 242

Kathleen Fraser is a Research Officer with the National Research Council Canada in the Text Analytics group. She works on various projects with collaborators within NRC, Canadian and international institutions, focusing on using computer algorithms to analyze and understand language. She is interested in applications of natural language processing for social good, through inter-disciplinary collaborations with experts in health care, psychology, education, and social science. Dr. Fraser has a B.Sc. in Physics from St. Francis Xavier University, M.Sc. in Computer Science from Dalhousie University and a Ph.D. in Computer Science from University of Toronto. Dr. Fraser is also an Adjunct Research Professor with the School of Computer Science, Carleton University.

9 Sessions

9.1 Session I

An Alignment-Free Approach to DNA Sequence Classification and Microbial Adaptation Analysis

by
Spandana Chereddy(UPEI),Gurjit S. Randhawa(UPEI),Kathleen A. Hill(UWO),Lila Kari(UW)
Supervisor(s):Gurjit S. Randhawa
Saturday, October 14, 8:10-8:30, MCDH 246

In the realm of genomics, accurate DNA sequence classification is essential for understanding evolutionary relationships, taxonomic distinctions, and microbial adaptations to different environments. This study uses ML-DSP, a groundbreaking method that combines supervised Machine Learning with Digital Signal Processing to create a versatile, alignment-free tool for classifying DNA sequences at various taxonomic levels. ML-DSP's innovation lies in its unique feature vector, which leverages the Pairwise Pearson Correlation Coefficient (PCC) between the magnitude spectra of the Discrete Fourier Transform (DFT) of a numerical representation of DNA sequences and those of other sequences in the training set. This approach not only enhances speed but also maintains an impressive average classification accuracy of over 97%. The methodology of ML-DSP involves three key components: DNA numerical representations, DFT analysis, and the computation of PCC for pairwise distances. Supervised Machine Learning classifiers are employed to classify new DNA sequences based on these distance measurements. A 10-fold cross-validation technique is used to evaluate classifier performance, ensuring the reliability of the results. Additionally, Classical Multidimensional Scaling (MDS) is used for visualization, creating a 3D representation of sequence relationships known as Molecular Distance Maps(MoDMap). To demonstrate ML-DSP's applicability, an extensive analysis of extremophilic microbial genomes was conducted. This dataset, comprising 693 high-quality genome assemblies of microorganisms adapted to extreme temperatures and pH levels, was categorized into two distinct groups: Temperature and pH. This categorization allowed for detailed taxonomic and environmental classifications. The results of the study showcase ML-DSP's proficiency in both supervised and unsupervised learning, exhibiting high accuracy in taxonomic classifications across various k-mer values. Furthermore, the method's ability to identify environmental components in genomic signatures of extremophiles suggests its potential as a valuable tool for understanding how microorganisms adapt to extreme conditions. In summary, ML-DSP represents a significant advancement in DNA sequence classification, offering a robust and efficient approach for researchers studying genetic diversity and microbial adaptation in various environments.

Analysis of racial and gender bias in MIMIC-IV clinical notes via natural language processing

by
Aakanksha Khandwaha (Mount Allison)
Supervisor(s):Dr. Amber Simpson (Simpson Lab at Queen's University)
Saturday, October 14, 1:30-1:50, MCDH 246

Artificial intelligence and machine learning methods are increasingly being used in healthcare settings for 'precision medicine', to develop better diagnostic tools, and automate administrative tasks. Recently, chatbots and medical large language models (LLMs) such as DxGPT (based on OpenAI's GPT-4 model) and Google's Med-PaLM-2 have started being deployed in hospitals, which are fine-tuned on electronic health record (EHR) data. However, bias towards patients based on their race, gender, and socioeconomic status is extremely prevalent in American and Canadian healthcare systems. For this reason, it is important to assess whether such biases are imbued within patients' clinical notes. One such source for EHR being the 'Medical Information Mart for Intensive Care' (MIMIC) datasets.

To quantitatively measure racial and gender bias in clinical notes, natural language processing (NLP) techniques were used on MIMIC-IV-Note, a cohort of deidentified free-text clinical notes for hospitalized patients. The insurance payer type was used as a proxy for socioeconomic status, and gender/self-reported race was provided for all patients in the dataset. Sentiment analysis was used to determine differences in positive and negative sentiment towards male and female patients, as well as white, Black, Asian, and Hispanic/Latino patients. Similar analysis is done for patients with Medicare/Medicaid and private insurance. The variation of commonly used words and topics between different genders and racial categories was also examined. It was found that there is a significant difference in sentiment between racial and gender subgroups. A limitation of this work is that it does not examine granular racial subgroups, as well as intersectional identities due to a lack of data on marginalized groups.

A Two Stream GNN for Point Cloud Classification

by

Muhammad Altaf Agowun (Saint Mary's)
Supervisor(s):Jiju Poovvancheri
Saturday, October 14, 8:30-8:50, MCDH 246

Point cloud classification has received tremendous attention in recent years with the adoption of LiDAR scanners and commodity RGB-D sensors for robotics, augmented reality, and urban planning applications. In this work, we explore the possibility of learning the global and local shape information through multiple topological representations of point clouds. To this end, we propose ”Node Spiked Reeb Graph (NSRG)” a modification to the standard Reeb graph that has potentials to capture the local as well as global features of the underlying shape. Aimed at capturing the surface geometry, we utilize alpha complex of the point cloud. A two-stream graph neural network (GNN) combining NSRG, and alpha complex has been designed and developed for point cloud classification. The proposed GNN has been compared and analyzed with the state-of-the-art. The performance of the proposed GNN has been evaluated using benchmark datasets such as ModelNet10, ModelNet40 and ScanObjectNN. The presentation will focus on the motivation and challenges of the point cloud classification, NSRG and alpha complex construction, GNN architecture and the experimental results.

Unveiling the Potential of Ubuntu Touch OS: Privacy-Centric Features, and Private Server Integration

by

Shivam Wadhwa (UPEI)
Supervisor(s): Cezar Câmpeanu (UPEI)
Saturday, October 14, 8:50-9:10, MCDH 246

In today's digital age, smartphones have seamlessly integrated into our daily routines, becoming an extension of ourselves from dawn till dusk. While smartphones offer unparalleled convenience, a persistent concern remains – the safeguarding of our personal data. Within the realm of smartphone operating systems, the duopoly of Android and iOS dominates. However, this study shifts the spotlight away from iOS, a proprietary system, and towards Android, an open-source platform largely governed by Google. A noteworthy contender in the open-source ecosystem is Ubuntu Touch OS, an OS that not only champions privacy and security but also remains comparatively lesser known.

This honours project aims to distinguish Ubuntu Touch OS from the widely embraced Android OS. The research endeavours to shed light on its implementation on smartphones and, more crucially, the creation of a private server to establish a seamless connection with smartphones running Ubuntu Touch OS. Extensive exploration of existing research reveals promising findings on the heightened security and privacy features of Ubuntu Touch OS, as well as its lightweight hardware demands.

Anticipated outcomes include a resource for researchers looking to delve into Ubuntu Touch OS implementation and a substantial contribution to the ever-evolving landscape of smartphone operating systems. This research is expected to elucidate the advantages of Ubuntu Touch OS while assessing its reliability in comparison to other operating systems. Moreover, by seamlessly integrating the desktop experience of Ubuntu with smartphones, this study hints at potential game-changing possibilities, particularly when coupled with advancements in artificial intelligence and machine learning.

A Submodelling-Based Approach to Expected Points in North American Football using Scikit-Learn

by

Baxter Madore (Saint Mary's)
Supervisor(s):Dr. Paul Muir(Saint Mary's)
Saturday, October 14, 9:10-9:30, MCDH 246

We apply the Python library scikit-learn to evaluate scoring probabilities and Expected Points in the National Football League. Expected Points (EP) is a metric common to many sports that quantifies the value of a given in-game situation in terms of the number of points that a team is expected to score or concede from that situation. We first create a model based on the widely-used nflScrapR EP model, and use a submodelling approach to separate data and handle it with one of several submodels trained only on relevant data for that particular scenario. This submodelling approach obtains significant accuracy improvements over using a single model in each considered scenario.

Saving energy for multitask soft real-time system using RL and DVFS

by

Xinyi Li (St. Francis Xavier)
Supervisor(s):Dr. Man Lin (St. Francis Xavier University)
Saturday, October 14, 9:30-9:50, MCDH 246

Energy consumption is an important indicator for evaluating soft real-time periodic system performance. This work is based on previous lab work of processing of single task. It aims to reduce energy consumption for periodic soft real-time systems with multi-tasks and multiple deadlines, by using Dynamic Voltage and Frequency Scaling (DVFS) and reinforcement learning to learn the complex task pattern, and then adjust the frequency based on the system environment. The method combines DVFS with reinforcement learning. DVFS is a commonly used method in the Linux kernel to reduce energy consumption where the DVFS governor in Linux decides how to adjust the frequency of the next time period according to the current load of the system, in order to save energy dynamically. The reinforcement learning method used in this work is Deep Recurrent Q-Network (DRQN). The general process of this method is to run a set multi-task multi-deadline workload. In the Linux kernel, neural networks are embedded inside the governor to compute the Q-value of each selectable frequency and the Q-value is used to select the next frequency. The embedded neural network consists of two parts, a layer of gated recurrent units (GRU) and a multilayer perceptron (MLP). The GRU output, along with other parameters are inputted into the MLP together. The frequency will also be put into MLP. Based on the Q value calculated by the MLP, the governor selects the frequency for the next time period. Each time the governor makes a decision, the inputs and outputs of the neural network, as well as other processing signals, are saved. After the whole workload is finished, this information is read by the user space to train the neural network with DRQN and update the parameters of the neural network. The updated parameters are sent into the kernel to continue running the workload for the next cycle. This process is repeated hundreds of times to learn a better neural network to save energy. A simple workload with 3 tasks and 3 deadlines is currently being tested for this system. The same CPU-intensive benchmark is started at times 0, 0.3, and 0.7, with corresponding deadlines of 0.3, 0.4, and 0.5. The workload was tested and found that the system can already recognize different deadlines and adjust the frequency so that each task is closer to its own deadline to save energy.

Verification of Heaps' law in large language model emulated text with implications for energy-efficient query processing

by Lai T Uyen (UPEI), Gurjit S. Randhawa (UPEI), Paul Sheridan (UPEI)
Supervisor(s):Paul Sheridan(UPEI), Gurjit S. Randhawa (UPEI)
Saturday, October 14, 9:50-10:10, MCDH 246

Heaps' law is an empirical law relating the number of unique words to the number of total words in corpora of textual documents. While this law has been validated in diverse human-authored text corpora, its applicability to large language model generated text remains unexplored. To addressing this gap, we investigate the validity of Heaps' law in GPT-Neo large language model emulated documents. In particular, we emulated corpora of PubMed abstracts using three distinct parameter sizes of the GPT-Neo model (125 million, 1.3 billion, and 2.7 billion parameters). Using simple linear regression, we validated Heaps' law on the GPT-Neo emulated corpora. Notably, as GPT-Neo model complexity increased, emulated vocabulary growth rate increasingly approached that of the human-authored PubMed abstracts. However, training GPT-Neo models with ever-increasing parameters did not yield linear improvements in adherence to Heaps' law as observed in the human-authored PubMed abstracts. This finding has practical implications for reducing the energy cost of large language model queries. For instance, we are presently investigating how to leverage Heap's law to minimize large language model energy consumption while still accurately mimicking human-authored text. This underscores the imperative for future advancements in large language models to prioritize energy-efficient strategies, possibly through architectural enhancements or optimization, over mere parameter inflation.

Fourier Transform-based Unsupervised Clustering of DNA Sequences

by
Wenwen Wang(UPEI),Gurjit S. Randhawa(UPEI),Kathleen A. Hill(UWO),Lila Kari(UW)
Supervisor(s):Gurjit S. Randhawa
Saturday, October 14, 10:10-10:30, MCDH 246

In the realm of genomics, the intricacies of data often challenge researchers, particularly when addressing the nuanced task of clustering less common or underrepresented DNA sequences. Our study introduces an advanced unsupervised learning framework, prominently integrating the capabilities of the Fast Fourier Transform (FFT). FFT, central to our methodology, enables the transformation of genomic sequences into a spectral format. This isn't merely a technical transition; it ushers in a fresh, insightful perspective for biological interpretation. With this spectral viewpoint, we can discern previously overlooked details and patterns within the sequences, shedding light on potentially crucial biological dynamics. Taking the advantage of FFT spectrum, the model is able to better utilize the subtle differences between data and also further help the data generation. Considering the sheer volume and complexity of genomic data, our approach further assimilates strategic dimension reduction techniques. These methods not only enhance computational speed but also complement the FFT-centric process, optimizing data management and ensuring a more effective unsupervised learning journey. In contemporary genomics, it's not uncommon for methods to lean on simulated or mimic-based sequences when faced with limited original data. In contrast, our strategy is rooted in the use of authentic genomic data, aiming to provide more genuine and relatable biological insights. Initial evaluations underscore the promise and adaptability of our framework. These preliminary findings not only validate its efficacy in clustering tasks but also suggest its potential as a reference point for future bioinformatics en-

deavours, bridging the gap between academic rigour and broader comprehensibility.

Enhancing Potato Crop Disease Detection and Classification Using Deep Learning and Robotic Imaging

by
Charanpreet Singh (UPEI), Gurjit S. Randhawa (UPEI), Aitazaz A. Farooque (UPEI)
Supervisor(s):Gurjit S. Randhawa (UPEI), Aitazaz A. Farooque (UPEI)
Saturday, October 14, 2:30-2:50, MCDH 246

Potato Virus Y (PVY) stands as a formidable and pervasive threat to global potato crops, inflicting substantial harm and presenting significant challenges. PVY, a member of the Potyvirus genus, is spread by aphids, contaminated seeds, and mechanical tools like pruners. Depending on the severity of the infection and the particular potato variety, yield losses attributed to PVY might range from 10% to a startling 80%. Several symptoms, including stunted growth, foliar mosaic patterns, leaf curling, and discoloration of both leaves and tubers, are brought on by PVY in potato plants. Some potato types suffer from severe foliar damage that ultimately results in plant death. Novel necrotic PVY strains have emerged, and they cause modest foliar signs that make it harder to identify infected plants. Notable PVY strains include the typical PVYO, which induces mosaic symptoms in most potato varieties, and the necrotic strains PVYNO and PVYNTN, resulting in vein necrosis in potato leaves. A pressing demand exists for innovative solutions aiding growers in pinpointing PVY-infected potato plants. Previous research has demonstrated the efficacy of deep learning-based convolutional neural networks (CNNs) in distinguishing between diverse plants, weeds, and diseases. This study further explores and expands upon the application of these models for PVY-infected plant detection. Various deep learning models are trained using imagery datasets comprising healthy and PVY-infected potato plants cultivated in greenhouse conditions. Optimal models are subsequently fine-tuned using field images collected by a custom-developed robot equipped with three RGB cameras, covering three rows of potato plants at a time. The robot employs an RTK system for geotagging captured images, enabling the precise localization of infected plants

in the field. Evaluation metrics encompass accuracy, precision, recall, and F1-score. Through 10-fold cross-validation, the trained models achieve classification accuracy scores exceeding 80% in discerning between healthy and PVY-infected potato plants. Significantly, the models demonstrate a capacity to reliably detect PVY-infected plants, even when symptoms are subtle, crucial for early intervention and curbing virus proliferation.

The Art of Contradiction Detection

by

Ethan Heavey(St. Francis Xavier), Milton King (St. Francis Xavier), James Hughes(St. Francis Xavier)

Supervisor(s):Milton King(St. Francis Xavier) and James Hughes(St. Francis Xavier)

Saturday, October 14, 1:50-2:10, MCDH 246

Natural language processing is a field of artificial intelligence that teaches computers how to analyze and understand human language. This research is commonly seen in applications that include topics such as machine translation, chatbots, and auto-fill software. A task within natural language processing, natural language inference, is a problem in which we assess relationships between sentence pairs (neutral, entailed, or contradictory) across three diverse datasets. We enhance models with two features: negation word count and persistent homology, the latter of which is a mathematical method for spatial pattern analysis. The best negation-augmented model, RoBERTa_Neg, improved classifier accuracy by 0.18%, surpassing baselines on two of three datasets. However, persistent homology vectors had a detrimental effect, reducing accuracy by 26.50% in the best-performing model (BBU_PH_SVM) and 60.20% in the worst-performing model (BBU_PH_MLP), falling short of established baselines. Combining these features in a single classifier showed interesting results but still resulted in reduced performance compared to using the negation feature alone. All results are compared and contrasted to previously established baselines on each dataset employed in this work.

Exploring the Genomic Signatures in a lower Dimensional Space

by
Riya Pandey (UPEI), Gurjit S. Randhawa (UPEI), Kathleen A. Hill (UWO), Lila Kari (UW)
Supervisor(s):Gurjit S. Randhawa
Saturday, October 14, 2:10-2:30, MCDH 246

The process of DNA classification involves organizing genetic material into taxonomic groups based on shared characteristics and evolutionary relationships. Chaos Game Representation (CGR) has been successfully utilized in conjunction with various Machine Learning algorithms to achieve high accuracies in DNA sequence classification. However, working with CGR-based feature vectors can be computationally demanding, leading to challenges in efficient classification due to the resulting high-dimensional feature space. Our goal is to identify a subset of CGR-based features through applying various dimensionality reduction techniques on the feature space while maintaining similar classification performance and enhancing computational efficiency. To assess the effectiveness of our approach, we employed a novel alignment-free supervised machine learning pipeline to classify over 15,500 complete mitochondrial genomes sourced from the NCBI reference dataset, spanning from the Kingdom to the Genus levels. Initially, we computed k-mer frequencies for the genomic sequences using CGRs, followed by the derivation of corresponding magnitude spectra by applying the Discrete Fourier Transform to the flattened CGRs. Subsequently, Machine learning algorithms were trained on the feature vectors generated by performing Principal Component Analysis (PCA) to the magnitude spectra. Results demonstrated that a subset of features identified through PCA achieved classification performance that was not only comparable but, in certain instances, even superior to that attained with the complete feature space. Furthermore, these selected features exhibited improved computational efficiency. Interestingly, we observed that even a few components, enough to explain 25% of the total variance, yielded accuracy levels equivalent to or surpassing those achieved with the entire feature space. In summary, this research highlights the potential of employing PCA for dimensionality reduction while preserving crucial features which enhances the efficiency and accuracy of the classification. Additionally, our approach

provided valuable insights into the genomic datasets by identifying the most significant features for classification.

A Solution to The Four-Dogs Pursuit Problem by a System of Ordinary Differential Equations and a Description of Instantaneous Position During the Process Based on the Solution

by

Mingyue Guo (UPEI)
Supervisor(s):TBA
Saturday, October 14, 2:30-2:50, MCDH 246

Abstract. In this article, we give a new solution of the Four-Dogs Pursuit Problem by a system of linear ordinary differential equations with constant coefficients. First, we set up the system of linear ordinary differential equations with constant coefficients based on the fact that the trajectories of the four dogs have nothing to do with their speed if the speed of them are equal to each other at arbitrary instantaneous time. Then, we worked out the trajectory of each dog by the eigenvalues and eigenvectors of the coefficient matrix of the system of linear ordinary differential equations. Then, we calculated the length of the trajectory along with the time needed for the whole process and described the instantaneous position of each dog during the pursuing process. Finally, the MATLAB Code of plotting the trajectories of the four dogs is given in the Appendix.

The Proper Spherically Symmetric Frame in Teleparallel Geometries

by

Hudson X. Forance (St. Francis Xavier)
Supervisor(s):Dr. Robert J. van den Hoogen
Saturday, October 14, 2:50-3:10, MCDH 246

Teleparallel gravity is an alternate approach to a theory of gravity, where instead of using curvature to describe the relationship between

matter, energy and space-time, we use torsion as the fundamental geometric quantity. In the last decade, $f(T)$-type gravity has been developed as a fully covariant theory that obeys local Lorentz invariance. We expand on former works by finding the local Lorentz transformation to a purely inertial frame for a class of teleparallel geometries, specifically spherically symmetric geometries.

We start with the spin connection for a general spherically symmetric teleparallel geometry and establish a system of matrix partial differential equations using the transformation laws for our spin connection. Solving these equations yields a transformation that can be applied to a generic spherically symmetric frame, bringing us to a purely inertial(proper) frame. Additionally, we explore sub-cases of the geometry that are relevant for describing cosmological effects.

An Extension of the Solow-Swan Growth Model

by

Timothy Power (Dalhousie) and Roman Smirnov(Dalhousie)
Supervisor(s):Roman Smirnov (Dalhousie)
Saturday, October 14, 3:10-3:30, MCDH 246

The aggregate Cobb-Douglas production function stands as a central element in the renowned Solow-Swan model in economics, providing a crucial theoretical framework for comprehending the determinants of economic growth. This model not only guides policymakers and economists but also influences their decisions, fostering sustainable and inclusive development. In this study, we utilize a one-input version of a recent generalization of the Cobb-Douglas production function proposed by Smirnov and Wang, thereby extending the Solow-Swan model to incorporate energy production as a factor. We offer a rationale for this extension and conduct a comprehensive analysis employing advanced mathematical tools to explore solutions to this new model. This approach enables us to gain insights into seamlessly integrating environmental concerns related to energy production into economic growth strategies, ensuring long-term sustainability.

9.2 Session II

The Science of Selection: Predicting NBA All-Stars Through Analytics

by Kevin Saville (St. Francis Xavier)
Supervisor(s): Kyran Cupido, Statistics Professor at StFX
Saturday, October 14, 8:10-8:30, MCDH 328

The purpose of this experiment is to compare a purely statistical approach to the current methodology used to select All Stars, which consists of voting from fans, media, players, and coaches. This study takes the statistics from the 2021-2022 NBA Season and uses them to predict the All Stars for the 2022-2023 season. All players who play more then a certain amount of minutes per game were selected as observations that fit the criteria mean, with their stats being compared to one another to determine the difference between All stars and non All stars. A linear discriminant analysis (LDA) was used to classify whether a player was an All star or not, based on 10 selected predictor variables. An LDA was used to help reduce the dimensionality of the data while preserving the information used to distinguish between classes, leading to a more accurate and efficient model for the study. Of the 27 players selected for the 2023 All Star game, the model predicted that 18 players would be selected, while the other 9 would not. The model also predicted 5 players that were not originally nominated to be All Stars would be considered using the model.

An Improved Algorithm for Gromovs Approximating Tree

by
Anders Cornect(Memorial)
Supervisor(s):Dr. Eduardo Martinez-Pedroza of Memorial University
Saturday, October 14, 8:30-8:50, MCDH 246

In 1987, Mikhail Gromov described an algorithm for constructing an approximating tree for an arbitrary finite metric space, with additive error proportional to a property of the space called the Gromov hyperbolicity. We discuss an improved version of this algorithm when the metric space is a graph, which runs in $O(n^2)$ time given the adjacency matrix of the graph. Joint work with Dr. Eduardo Martinez-Pedroza.

Proper Circular Arc Graphs and Circular Robinson Spaces

by

Ethan Saunders (Dalhousie)
Supervisor(s):Dr. Jeannette Janssen, Dalhousie University
Saturday, October 14, 8:50-9:10, MCDH 328

A Graph G is called a proper interval graph if it is the intersection graph of a set of intervals on the real line with no interval properly containing any other. G is called a proper circular arc graph if it is the intersection graph of a family of arcs on a circle with no arc properly containing any other A similarity space is a set X along with a similarity measure s, where $s(a,b) = s(b,a)$ is the similarity between elements a and b in X. A similarity space (X,s) is called a circular Robinson space if there exists a circular ordering that is consistent with the similarity measure. Applications of PCA and proper interval graph recognition include scheduling, traffic control, and DNA mapping. Applications of Circular Robinson space recognition include musicology, planar tomographic reconstruction, and hypertext orderings. I will present an algorithm for recognizing PCA graphs. The algorithm builds on a classic algorithm for recognizing proper interval graphs. I will also present a result stating that a binary similarity space (one where every similarity is either 1 or 0) is circular Robinson if and only if its similarity matrix is the adjacency matrix of a PCA graph. I will discuss how this result could be applied to Circular Robinson Space Recognition.

Reducing the T-count of quantum C^n-NOT gates

by

Andrew Allen (Dalhousie)
Supervisor(s):Neil J. Ross, PhD. Dalhousie University,Peter Selinger, PhD. Dalhousie University,Andre Kornell, PhD. Dalhousie University,Frank Fu, PhD. Dalhousie University
Saturday, October 14, 2:10-2:30, MCDH 329

C^n-NOT gates are important for realizing classical algorithms in a quantum context. To construct circuits for these gates in a fault-tolerant fashion, it is of interest to minimize the T-count of these gates, as the T gate is costly to physically implement compared to the Cliffords. I review optimizations made to the T-count of the C^n-NOT gate family over the past decade — in particular, reductions of the T-counts for the Toffoli and C^3-NOT from 7 to 4 and 8 to 6, respectively, and implications for the general C^n-NOT count. This brings us to the current state-of-the-art count of $4n-6$ for n controls. Improvements past this current count have proved elusive, so I consider other potential approaches to this problem. ZX calculus has been used to optimize C^n-NOT, so this area could be a candidate. Lower bounds on the T-count could also be investigated for insight into how much the gates can be further optimized.

QFASApp: A Friendly Interface for QFASA

by

Anson Green (UNB Saint John)
Supervisor(s):Dr. Connie Stewart (UNB Saint John)
Saturday, October 14, 9:30-9:50, MCDH 328

When studying the behavior and diets of marine predators, there are many limitations to data collection that can make it difficult to generalize results. Systems that allow for direct observation of feeding are typically limited either to a very specific region or by financial factors. Many indirect diet estimation methods, such as stomach content analysis, are lethal and can be biased towards the last meal and the recording of prey with hard parts that take longer to digest. Quantitative Fatty Acid Signature Analysis (or QFASA) is a statistical model that ameliorates these concerns, using proportions of fatty acids stored in the adipose tissue of a marine predator's body to estimate diet over longer periods of time without killing or harming predators during data collection. The QFASA model requires a database of fatty acid signatures for all potential prey in the predator's diet and attempts to match the fatty acid signature of the predator to the prey. Since the QFASA model is used by biologists worldwide and is regularly updated with new functionality for model assessment and inference, a QFASA R package was developed. However, the R package still requires a multi-step process of data formatting and manipulation to function,

a process that can be intimidating for users without a background in R. This project aims to create a user-friendly interface through the Shiny package in R that receives a set of simpler inputs and runs the QFASA process in the background, displaying a boxplot of the QFASA output. This would allow users new to R to run QFASA, as well as reducing the tedium of running the process repeatedly for statisticians to observe changes in the data after changing inputs.

Submission Ideas in Orlicz Spaces

by

Griffin Bartlett (Cape Breton)
Supervisor(s):Scott Rodney (CBU)
Saturday, October 14, 9:50-10:10, MCDH 328

In this talk I will discuss some research I did over the summer as an NSERC USRA research assistant at CBU. I will introduce the ideas and basic concepts in Orlicz Function Spaces. Following this, we will investigate the optimality of constants when comparing Orlicz norms associated to comparable Young functions.

Modelling the Effectiveness of Non-Pharmaceutical Interventions on Influenza A & B

by

Heather MacTavish (Mount Allison)
Supervisor(s):Dr. Betti, Mount Allison
Saturday, October 14, 10:10-10:30, MCDH 328

Non-pharmaceutical interventions such as masking and lockdowns that were used as a COVID-19 control also acted as an effective way to limit the spread of other endemic viruses. Using data from both historical and current influenza A & B seasons as well as COVID-19 data, we use SIR and SEIR models to show the effect that non-pharmaceutical interventions specifically used to target COVID-19 had on the seasonal influenza virus. Under the assumption that the effectiveness of non-pharmaceutical interventions remains constant in different viruses, we can explore how the observable effects of these NPIs on endemic infections can provide insight into how effective they may be and what the ”true reproduction numbers” are for other infectious diseases that may emerge in populations.

Domination Polynomials of Graph Operations

by

Erin Hughes (Acadia)
Supervisor(s):Dr. Iain Beaton
Saturday, October 14, 1:30-1:50, MCDH 328

This talk focuses on the general forms of directed domination polynomial of graphs. The domination polynomial enumerates all the dominating sets of a graph G. The dominating sets of a graph are sets such that all vertices are either in the set or adjacent to a vertex in the set. Additionally, all the graphs investigated in this paper are directed. In this talk we will investigate the domination polynomials of graphs having undergone the join or corona operation. This work was done in collaboration with my supervisor, Dr. Iain Beaton of Acadia University.

Misère Cricket Pitch

by

Ethan Saunders (Dalhousie)
Supervisor(s):Dr. Richard Nowakowski, Dalhousie University
Saturday, October 14, 1:50-2:10, MCDH 328

Cricket Pitch is a combinatorial game whose result is determined by who makes the last move. It was created in 2010 by Nowakowski and Ottaway, as an example of a new class of games. They solved the normal play version – where the last player to move wins – and left the misere version – where the last player to move loses – as an open problem. I will present an overview of the solution to the misere version originating from a games workshop held at Dalhousie over the summer. The main future goal is to be able to analyze disjunctive sums of misere cricket pitch games. A disjunctive sum of games is a set of games in which each player can choose to play a move in any one of the games in the set on each turn. The full analysis would involve finding an algebra that allows us to quantify the difference between positions. The new class has the property that any position that a player can reach with 2 or more consecutive moves can be reached by that same player with just one move. These are option closed. Finding an algebra that quantifies the difference between misere option closed games is an important open problem and a solution to multi-board cricket pitch could help to better understand option closed games in general.

All oriented paths and cycles are unimodal

by
Nimarjeet Bajwa (Acadia)
Supervisor(s):Iain Beaton (Acadia University)
Saturday, October 14, 2:10-2:30, MCDH 328

A polynomial is said to be unimodal if its coefficients are non-decreasing and then non-increasing. The domination polynomial of a graph G of order n is the polynomial $D(G,x) = \sum_{i=\gamma(G)}^{n} d(G,i) \cdot x^i$, where $d(G,i)$ is the number of dominating sets of G of size i, and $\gamma(G)$ is the domination number of G. In this presentation, we will show that the directed domination polynomial of all oriented paths and cycles are unimodal.

Mixing Models for Domination Reconfiguration

by Logan Pipes (Mount Allison)
Supervisor(s):Dr. Margaret-Ellen Messinger, Mount Allison University
Saturday, October 14, 2:30-2:50, MCDH 328

A dominating set S of a graph G is a subset of its vertices such that every vertex of G is either an element of S, or adjacent to an element of S. For a graph G, consider all of its dominating sets. We can construct a graph whose vertices are these dominating sets, where two dominating sets are adjacent if one can be transformed into the other via a predetermined rule. This graph is called the (domination) reconfiguration graph of G. Many natural questions arise concerning the structure of these graphs, including questions of connectivity, diameter, and Hamiltonicity, to name a few. Various rulesets for adjacency are introduced, with some corresponding answers to these questions. A new, mixed model is proposed, along with some preliminary results.

Maximum Nim with a Pass using Fractional Restricions

by
Samuel Sarria Hurtado (UPEI)
Supervisor(s): Dr. David Horrocks(UPEI)
Saturday, October 14, 2:50-3:10, MCDH 328

Combinatorial game theory studies games that are deterministic, have two players, each player has perfect information about the game, and the game will end eventually. In this field of mathematics, there are partial games and impartial games. Partial games are those where the move-set for one player is different from that of the other player, and impartial games are those where the move-sets are the same. Nim is a classic impartial game in the field of combinatorial game theory. We set out to study one of its variants, namely, maximum Nim. In this variant, a player may at most take a fraction of a pile of their choosing rounded up, and at least one stone (unless they use the pass move). In addition to having this maximum restriction, we also consider a pass move that can be played only once in a game, and we provide a relationship between maximum Nim with a pass and without a pass. In the analysis below, we first set out to find the winning positions for these games with certain fractions as maximum restrictions, namely those of the form $\frac{l-1}{l}$ (where l is a positive integer) and then for the rest. The winning positions can be determined via integer sequences and recurrence relations. Beyond the winning positions is the Sprague-Grundy integer function, which assigns a positive integer to a move. Once the winning positions have been determined via the aforementioned methods, the Sprague-Grundy value for any pile size can be determined similarly. Finally, we study the game with multiple piles of stones, relying on the important theorem that the Sprague-Grundy value of a game is the nim-sum of the individual piles involved, with and without a pass.

Investigating the Spatial Distribution of the Underlying Risk Factors for Covid-19 in Canada

by
Simon Maltby (St. Francis Xavier)
Supervisor(s):Dr. Kyran Cupido (St. Francis Xavier)
Saturday, October 14, 3:10-3:30, MCDH 328

While the impact of Covid-19 in Canada has been well documented at the provincial and national levels, it is not fully understood how the severity of the impact of the pandemic varied locally across space. As such, a better understanding of the spatial distribution of the risk factors for Covid-19 would provide valuable information for predicting potential hot spots of the pandemic in Canada at a regional level. The primary objective of this research project was to detect patterns in the spatial distribution of the underlying risk factors for Covid-19 across the census divisions of Canada. After identifying the risk factors of interest, a spatial weights matrix was constructed, and the Moran's I coefficient was used to determine hot spots and cold spots for these variables across Canada. Then, principal components and model-based clustering were employed to pool together data and identify which regions were closest to each other in terms of risk for Covid-19. The defining characteristics of each cluster and how their frequency varied across Canada were also determined. Upon the release of more refined Covid-19 data, the findings of this study could be compared to how the pandemic affected different regions at the census division level.

9.3 Session III

Well-being of Blind or Low Vision Canadian Adolescents

by

Aadesh Nunkoo (UPEI)
Supervisor(s):Dr. Michael McIsaac (UPEI)
Saturday, October 14, 8:10-8:30, MCDH 329

Our research investigated well-being outcomes including life satisfaction, mental health, and loneliness of (1) Canadian blind and low-vision adolescents and (2) Canadian blind and low-vision adolescents with multiple disabilities, compared to non-disabled adolescents. We analyzed cross-sectional data from the 2017-2018 Canadian Health Behavior in School-aged Children (HBSC) survey. Our sample included 19268 Canadian adolescents comprising of 18837 non-disabled adolescents and 431 blind or low vision adolescents. We employed Chi-square tests (with corresponding 95% confidence intervals) to assess differences in proportions pertaining to the well-being outcomes between these two groups. Our findings revealed significant differences in subjective and psychological well-being between the two groups ($p < 0.01$). These disparities are further made worse for Canadian blind and low vision adolescents with additional disabilities.

Dualizability through 2-categories

by
Daniel Teixeira (Dalhousie)
Supervisor(s):Theo Johnson-Freyd (Dalhousie)
Saturday, October 14, 8:30-8:50, MCDH 329

In higher category theory, the concept of full dualizability can be, perhaps surprisingly, encoded as a sequence of 2-categories whose adjacent entries overlap in a certain way. The goal of this project is to understand these ladders categorically. In this talk we will build up to the definition of full dualizability and ladders.

Causal Effect of Factors Impacting Participation in Music Education in Canadian Adolescents

by Saurav Neupane (UPEI)
Supervisor(s):Dr. Michael A. McIsaac (UPEI)
Saturday, October 14, 8:50-9:10, MCDH 329

Music education has many positive outcomes, including better academic achievement and cognitive development. Despite the benefits, less than 50% of secondary students participate in in-school music education in Canada. If in-school music education benefits an adult's development, unequal access might hamper their development. School demographics, funding, and resources are some of the factors affecting equal access to music education in Canadian adolescents. Our study aims to find the causal effects of social location variables such as Gender, Socio-Economic Status (SES), and Urban-Rural Status on arts participation in Canadian adolescents using the Health Behaviour in School-Aged Children (HBSC) survey. Our study found that female students had higher odds of arts participation than male students, students with higher SES had higher odds of arts participation than students with lower SES, and students in large population centers had higher odds of arts participation than students in rural areas. These results provide valuable insights into the factors contributing to Canada's unequal access to arts and music participation.

Prediction models of mortality for people with cystic fibrosis

by
Janet Cheng (Dalhousie)
Supervisor(s): Dr. Orla Murphy(Dalhousie), Dr. Sanja Stanojevic (Dalhousie)
Saturday, October 14, 9:10-9:30, MCDH 329

Cystic fibrosis (CF) is a rare genetic disease that causes lung damage and early death. Lung transplantation is a treatment option that prolongs life for people with end-stage CF. Due to the scarcity of donor lungs, transplant candidates need to be carefully selected so that their long-term survival is maximized. The timing of transplantation is also crucial for survival. Accurate prediction of who has an increased risk of death can be an aid in making decisions about who should be prioritized in getting the transplant and what the optimal timing is. I conducted a comprehensive literature search to identify 16 studies that have developed mortality prediction models for CF. These studies developed different prediction models with varying sets of variables and employed a variety of variable selection methods. A seminal paper by Kerem, et al., published in 1992, showed that FEV1 (i.e., forced expiratory volume in one second, a measure of lung function) predicts mortality in individuals with CF. This result has been used for establishing consensus guidelines for selecting lung transplant candidates, which are still in use today. The studies that followed Kerem, et al., have yet to be applied clinically. To evaluate the studies, I compared them against the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) Statement checklist. Overall, reporting transparency was poor across the 16 studies, especially for model performance metrics. These 16 studies also have varying outcomes, sample sizes, age, population, and data sources. These factors make direct comparison between the studies challenging and preclude the possibility of identifying an optimal prediction model. Therefore, future analyses will be necessary for comparing the models and for finding an optimal model. By fitting the models on

the Canadian CF registry dataset and clearly assessing them using TRIPOD, I will provide a systematic summary of the models and statistical tools available to develop prediction models for people with CF. This work will inform decision-makers to help identify individuals who would benefit from transplant referrals.

Identifying Gambling Personae Through Machine Learning Algorithms

by
Éloïse Soucy (Université de Moncton)
Supervisor(s):Salah El Adlouni (Université de Moncton), Sophie Léger Auffrey (Université de Moncton),Jalila Jbilou (Université de Moncton et Centre de formation médicale du Nouveau-Brunswick)
Saturday, October 14, 9:30-9:50, MCDH 329

There is a growing need for gambling disorder studies. Especially with the easy access of online gaming which made gambling more accessible than ever. This research aims to identify distinct gambling personae through machine learning algorithms and could potentially lead to a model that could detect potential pathological gamblers in the early stage by identifying the playstyle.

This work has a review of related theories and used gambling sessions data as an application for these methods. We used Uniform Manifold Approximation and Projection(UMAP) on the data as a dimension reduction, and did the clustering with k-means. Lastly, we used SVM on the data as a classification algorithm

A Comparative Study on Optimization Methods for Maritime Transport Routing

by
Louis Bu (Dalhousie)
Supervisor(s):Dr. Robert Milson and Dr. David Gómez-Ullate
Saturday, October 14, 9:50-10:10, MCDH 329

Maritime transport of cargo is the most cost-efficient method of transportation. However, it accounts for near 2% of global greenhouse gas

(GHG) emissions. In recent years, we have seen initiatives set out by the International Maritime Organization (IMO) and the United Nations to reduce and ultimately eliminate the marine transport's carbon footprint as an effort to limit climate change. In 2018, IMO set out a road plan, titled *'The IMO Initial Strategy for Reducing Greenhouse Gas(GHG) Emissions, and its Follow-up Actions Towards 2050'*, and again updated in July 2023 with an objective of reducing emissions by at least 20% by 2030 and at least 70% by 2040, compared to levels 2008.

The IMO also outlined different approaches to achieve the abovementioned goals, such as improving the Energy Performance Index of existing ships, developing a speed optimization mechanism, developing carbon-neutral fuel sources, and alternative propulsion methods. Projects such as fitting sails and utilize wind energy, however this requires long development time and retrofitting the current fleet.

An immediate, and relatively cheap class of solutions that can help achieve such goal set out by the IMO is called *Weather Routing*. Weather Routing is an umbrella term for the optimization algorithms that utilize current/forecasted meteorological data (e.g., winds, currents, etc.) to our advantage to reduce fuel consumption and emissions. We can also use meteorological data in conjunction with other data sources to identify and avoid pirate activity zones, ecological sensitive and protected zones, and adverse weather conditions. With a well-designed and well-optimized Weather Routing algorithm, we can save up to 30% of fuel consumption on a single voyage, which can reduce the fuel costs for shipping companies which in turn reduce carbon emissions and reduce the carbon footprint of the maritime transport sector, it can also improve voyage safety and secure both sailors on board, the ship, and the cargo.

In this talk we propose a class of optimization methods of 3 different solutions: *Hybrid Search*, *A^* Search*, and *CMA-ES Genetic Algorithm*, and compare their performances with real world metrics. Each of these solutions approaches the problem from a different area of mathematics, insights from differential equations, graph theory, and evolutionary genetic algorithm respectively. We will discuss the principles behind them, and validaty of results.

Spatiotemporal Models for Exploring Variability in Scallop Condition across the Bay of Fundy

by
Joy Liu (Dalhousie)
Supervisor(s):Joanna Mils Flemming (Dal),Orla A. Murphy (Dal)
Saturday, October 14, 9:10-9:30, MCDH 328

Sea scallops (Placopecten magellanicus) comprise the fifth largest fishery in Canada, the vast majority of which occurs in the Maritimes. The target of the fishery is the adductor muscle (i.e., the meat) with size varying both in time and space. We use spatiotemporal models that incorporate environmental data (depth, salinity, etc.) to explore the variability in scallop meat weight (and scallop shell height) in the Bay of Fundy. These models elicit some interesting patterns regarding the condition of scallops as well as improve current estimates of scallop meat weight for this area. These results can improve science advice and lead to more sustainable management of the stock.

Quaternionic methods in exact synthesis

by
Linh Dinh (Dalhousie)
Supervisor(s):Dr. Julien Ross (Dalhousie University)
Saturday, October 14, 10:10-10:30, MCDH 329

The exact synthesis problem is an important problem in the theory of quantum computing, where one breaks down a given operator into simpler ones to maximize efficiency and reduce cost. Viewing operators as matrices, the problem becomes a matrix factorization problem. In this talk, I will introduce and compare two methods for the exact synthesis of 2x2 unitary matrices, focusing on the case of matrices with entries in the ring $\mathbb{Z}[1/2, \omega]$, where omega is an eighth root of unity. The first method works by considering the columns of the matrix one by one, while the second method considers the matrix globally by interpreting it as a quaternion. Afterwards, I will discuss possible extensions of both methods, and some of the difficulties one might face in generalizing them.

Exploring the Impact of Bubble Strategies on the Spread of Infectious Diseases

by
Jingyu Li (Dalhousie)
Supervisor(s):Dr. Edward Susko(Dalhousie), Dr. Lam Ho(Dalhousie)
Saturday, October 14, 1:30-1:50, MCDH 329

The emergence of the COVID-19 pandemic has generated a great deal of scientific interest in strategies to effectively respond to epidemics. Compartmental models, such as the Susceptible-Exposed-Infected-Removed (SEIR) model, are popular tools that allow us to study the behaviour of various epidemics. However, the model assumes that the population is well-mixed and does not take into account public health policies such as social distance and bubble strategies. Here, a Bubble SEIR model is constructed based on the stochastic continuous-time Markov chain SEIR model. The Bubble SEIR model divides the population into subpopulations called bubbles. The model assumes that each subpopulation is well-mixed and also allows the possibility of movements between bubbles. Simulations using the developed model were conducted to investigate the effect of the bubble strategies on the spread of infectious diseases. In the simulation study, we will discuss the effects of parameters of the model including rate of movement between bubbles, basic reproduction number, latency period, population size and initial number of infecteds, on the final size and the duration of the epidemic. The simulation experiments show that the bubble strategies are effective in controlling the spread of infectious diseases. When implementing the bubble strategies, achieving zero contact between bubbles can be challenging. However, curbing the rate at which individuals move between bubbles can have a positive impact. Furthermore, while changes in infection and removed rates affect the final size of the epidemic, changes in the basic reproduction number have a much larger impact. While the latency period varies depending on the nature of the virus, it is worth noting that the bubble strategies tend to be more effective in controlling the spread of epidemics characterized by shorter latency period. More bubbles proved to be more effective in controlling epidemics when the population size increased, and also demonstrated that bubble strategies should be implemented early.

Modelling Virus Containment with Self-disseminating Vaccines on n-Dimensional Grids

by

Dylan Pearson(Dalhousie), Margaret-Ellen Messinger(Mount Allison), Melissa Huggan(Vancouver Island University)

Supervisor(s): Jeannette Janssen, Dalhousie and Nancy Clarke, Acadia

Saturday, October 13, 1:50-2:10, MCDH 329

Controlling the spread of contagion in wildlife is a prominent immunological challenge. One method for doing this is to employ the use of self-disseminating vaccines: vaccines that are capable of spreading autonomously. We study the most efficient distributions of these vaccines using graphs. We determine the minimum number of self-disseminating vaccines required to contain the spread of a virus on the infinite n-dimensional Cartesian and strong grids, and include strategies to do so as quickly as possible

Copula Mixture Regression Models for Multivariate Response Data

by

Claire Cui(Dalhousie), Orla A. Murphy(Dalhousie), Paul D. McNicholas(McMaster)

Supervisor(s): Prof. Orla A. Murphy(Dalhousie),Prof. Paul D. McNicholas(McMaster)

Saturday, October 14, 2:30-2:50, MCDH 329

Clustering, an unsupervised method in the realm of data analysis, serves as a powerful tool for uncovering hidden patterns and structures within complex datasets. In recent years, the use of mixtures of multiple linear regression models in clustering has gained popularity due to its ability to account for underlying heterogeneity in the data and provide a more representative interpretation of covariate effects. However, there is a paucity of these models for multivariate response cases, particularly when dealing with dependent responses. One approach that has been applied in the case of multivariate response data is copula regression models. Copulas can be seen as representing the dependence structure of a random vector and are joint distribution functions with uniform margins. In copula regression, a copula function is employed to induce dependence between different response variables through the

random error term in the regression model. The concept behind using copula regression models is that they allow us to capture complex dependencies among variables while still maintaining flexibility and interpretability. By incorporating copulas into our analysis, we can better understand how different covariates affect each component of the multivariate response.

In this work, we propose a copula-based finite mixture of regression (CMixR) model for clustering and interpreting covariate effects in heterogeneous multivariate response data and present an ECM algorithm for estimation. The model performance is tested using a simulation study and through data analysis on the morphological properties of purple rock crabs. The results obtained from this model are shown to give excellent clustering performance, as evidenced by the high adjusted Rand index (ARI) values.

Visualizing Qudit Controls With Sheets

by

Scott Wesley (Dalhousie)
Supervisor(s): Neil (Julien) Ross (Dalhousie)
Saturday, October 14, 2:50-3:10, MCDH 329

Graphical methods have a long history in mathematics, dating back to Euclid's Elements. However, this tradition was abandoned in favour of Hilbert's program. More recently, research in topology and categorical algebra have driven a resurgence in the development of graphical techniques. Of particular interest are the two-dimensional string diagrams of categorical algebra, which share deep connections with quantum computation. This paper reviews some important categorical concepts in the theory of quantum circuit diagrams, and describes an ongoing effort to study controlled qudit operations through the lens of three-dimension sheet diagrams.

Validation of Earth System Models

by

Ethan O'Connell (Dalhousie)
Supervisor(s):Dr. Michael Dowd(Dalhousie)
Saturday, October 14, 3:10-3:30, MCDH 329

Earth system models are numerical models that simulate several aspects of the Earth system. They use coupled systems of partial differential equations to model the fluid dynamics of the atmosphere and the oceans, as well as other chemical and biological processes. With increases in computational power and model complexity, as well as the availability of calibration datasets, there is an increased need for accurate model assessment to inform decision making. In this paper we provide an overview of common statistical methods and protocols for the validation of Earth system models. We present some popular numerical metrics and summary diagrams used for the qualitative and quantitative evaluation of model performance. An example procedure is demonstrated and applied to the simulation of nearshore ocean temperatures in the Scotian Shelf.