ARIA

APPLIED RESEARCH IN ACTION

2020 RESEARCH INTERNSHIP PROJECTS

Computer Science
UNIVERSITY OF TORONTO

MScAC Master of Science
in Applied Computing

THANK YOU
TO OUR INDUSTRY PARTNERS

Agnostiq

Amazon

Bridge7 Oncology

CPP Investments

Deloitte

Ecobee

Ethoca Technologies Inc.

Geotab

Google

Kensington Vision and
Research Centre

ICICI Bank

Integrate.ai

Layer 6 AI

League Inc.

Loblaw Digital

Mila

ModiFace Inc.

NVIDIA

Okta

Pearson Canada Inc.

Rakuten Kobo Inc.

RSVP.ai

Samsung AI Centre

Scotiabank

SOTI Inc.

Surgical Safety Technologies

Synex Medical

Tactual Labs Co.

Tealbook

Vanguard Investments Canada

VerticalScope Inc.

ARIA 2020 commemorates the beginning of the Master of Science in Applied Computing (MScAC) Program's ten-year history — a decade of demonstrated excellence that has seen our students go on to become technological leaders in Toronto and beyond. Our annual Applied Research in Action showcase marks the capstone of our students' successes, showcasing marvelous innovations they have developed with their industry partners. We are proud to recognize this year's MScAC student cohort and showcase their array of cutting-edge projects.

This year, for the first time, ARIA was a completely virtual event and we broke all attendance records. With participants joining us from all over the world, we welcomed them to celebrate this milestone achievement and to discover the latest innovations in technology. We were particularly pleased to welcome students from many countries who came to learn more about our program, with many subsequently choosing to apply. Our sincerest thanks to everyone who made ARIA 2020 possible and who make the MScAC Program one of the world's best Applied Computing programs.

Congratulations to our 2019/20 cohort and welcome to our MScAC alumni family!

Arvind Gupta
Professor
Academic Director, Professional Programs
Department of Computer Science

THE ARIA AWARDS
2020 WINNERS

Abenezer Teklemariam

MScAC 2019 Cohort

STUDENT INNOVATION AWARD

Nominated by their industry supervisor, this award is presented to a student in the internship phase of the MScAC program who has demonstrated innovation, dedication and creativity in the completion of their internship project.

Shi Yu

Vanguard Investments Canada

INDUSTRY ICON AWARD

The Industry Icon award is a student-nominated award, presented to an industry partner who has taken on the role of an MScAC industry supervisor and demonstrated exceptional commitment, time, mentorship and skilled attention to developing the talents of their students.

Fanny Chevalier

Department of Computer Science,

University of Toronto

FACULTY RECOGNITION AWARD

The Faculty Recognition Award is a student-nominated award, presented to a faculty supervisor who has demonstrated exceptional commitment, time and skilled attention to developing the talents of their students.

Geoffrey Peddle

MScAC 2010 Cohort

ALUMNI OF THE YEAR AWARD

The Alumni of the Year Award is presented to a member/members of the MScAC alumni network who have made significant contributions to the development of the program.

Cara Alexander
Faisal Habib
Shakti Kumar

MScAC 2019 Cohort

STUDENT PRESENTATION AWARD

The Student Presentation Award is awarded to the student(s) who is considered to have the best presentation at the ARIA showcase.

Eugene Fiume

Director, MScAC Program (2010-2016)

MSCAC SPOTLIGHT AWARD

The MScAC Spotlight Award is presented to an individual who has made a tremendous impact on the program.

2020 RESEARCH INTERNSHIP PROJECTS

Quantifying Degree of Obfuscation in Quantum Circuits with Graph Neural Networks

Company: Agnostiq
Student: Cara Alexander
Industry Supervisors: Oktay Goktas, Edwin Tham
Academic Supervisor: Henry Yuen

Numerous industrial applications of quantum computing are constrained by privacy sensitivity, so in the absence of new security solutions, organizations working with confidential information are unable to engage with quantum hardware. Agnostiq has developed software to obfuscate quantum circuits, enabling users to experiment with quantum processors over the cloud while protecting not only their sensitive data, but also the algorithms themselves. This project focuses on developing heuristics to quantify the degree of obfuscation achieved by this software. Definitions of cryptographic obfuscation including black-box, best-possible, or indistinguishability, are either provably impossible or contextually infeasible. Hence there is a need for new metrics through which users can compare different levels of obfuscation so that average-case security can be balanced with computational overhead.

This work utilizes an information-theory based conceptualization of how much entropy is injected into the circuit by the obfuscator to quantify how well the software hides sensitive gate parameters. Graph neural networks are used as circuit type classifiers to assess the degree to which the connective structure of a circuit is obscured. A distinctive behavior was observed wherein the obfuscator's ability to hide the circuit structure from the neural network increases dramatically around the point where 1% of a circuit's two qubit gates are added by the security software. Adding additional obfuscation beyond this drop-off does not appear to provide any additional gain in confusing the classifier. The future of this research will involve incorporating these heuristics into a unified framework to score the degree of obfuscation.

Project Natural Hand

Company: Tactual Labs Co.
Student: Jianda Chen
Industry Supervisors: Bruno de Araujo, Valkyrie Savage
Academic Supervisor: Fanny Chevalier

This project explores 3D pose hand reconstruction using machine learning models, with sensor data from innovative input devices. The goal was to propose an approach that provides real-time high fidelity hand reconstruction and understands how users perceive its quality, in order to improve user experience and social interaction in AR/VR. Our methodology consists of training a deep learning network, which learns the mapping from sensor signals to hand pose joint locations. We assess the quality of predicted hand motions with user studies. The training data is collected using both the Tactual Lab wearable device and the leap motion camera. At the current stage, we have implemented a prediction pipeline and a visualization script that shows the prediction of the user's hand. New evaluation metrics are being explored with respect to user perception ratings in different application scenarios. In the next steps, we will perform a new data collection using a motion capture system, refine prediction methods and generalize the model to work with multiple users and wristband placements.

Speeding up Landmark Detection by Network Architecture Search

Company: ModiFace Inc.
Student: Zihao Chen
Industry Supervisor: Soheil Seyfaie
Academic Supervisor: Anthony Bonner

Makeup simulation is a crucial Augmented Reality (AR) task in the beauty industry. Currently, handcrafted convolutional neural networks (CNN) are applied to localize a user's face landmark positions for the simulation. These models have state-of-the-art performances, but they are not light-weighted enough to run smoothly on webpages. A faster and more accurate CNN model can significantly enhance the user experience, yet very challenging for a human to finetune.

In this project, instead of manual design, our goal is to use Neural Architecture Search (NAS) to find landmark detection models with better capacity, accuracy, and efficiency than the current state-of-the-art. With a search space of 8×10^{15} different candidates, an evolutionary search is conducted to explore the optimal model structures efficiently. Under these proper search space and search strategy, the selected model has a time cost 33% lower than the baseline while maintaining accuracy. This NAS approach will be customized to different platforms for further improvement.

Exploring Fluent Query Reformulations with Text-to-text Transformer and Reinforcement Learning

Company: Vanguard Investments Canada
Student: Jerry Zikun Chen
Industry Supervisors: Haoran Wang, Shi Yu
Academic Supervisor: Scott Sanner

Query reformulation aims to alter noisy or ambiguous text sequences into coherent ones closer to natural questions. This is to prevent errors from propagating in a client-facing Natural Language Processing (NLP) pipeline and promote better communication with users. Besides, it is crucial to maintain performance in downstream environments like question answering when rephrased queries are given as input. We show that under the previous framework (AQA), attempts to alter reinforcement learning (RL) algorithms do not bring significant benefits to either reward acquisition or sequence fluency. Instead, we leverage a query-reformulating text-to-text transformer (QRT5) and apply policy-based RL algorithms to further nudge this reformulator and obtain better answers downstream by generating reward-acquiring query trajectories. QRT5 shows better sample efficiency in RL to achieve the same level of QA performance as the previous approach. It can generate reformulations with more readability, based on query well-formedness evaluations and can generalize to out-of-sample data. Our framework is demonstrated to be flexible, allowing reward signals to be sourced from different downstream environments such as intent classification.

Embedding Representation Based Search Personalization

Company: Loblaw Digital
Student: Zhou Fang
Industry Supervisor: Richard Downe
Academic Supervisor: Roger Grosse

In online shopping, search results often have inherent ambiguity. Two customers using the same term as the search query might have completely different expectations for the search results. Therefore, a globally trained ranking model, which is optimized for a generic notion of relevance, is unable to learn the specific user preferences. To provide a better online shopping experience at Loblaws, we aim to enable the ranking model to learn from user behaviors and then deliver search results that are personalized to all of them.

In this project, we designed and implemented a two-stage ranking model. In the first stage, the candidate items retrieved by Elasticsearch are first ranked by personalized popularity. In the second stage, we leverage embedding representations of user and item to infer user-item affinity. Specifically, we first train item embeddings so that closely related items in the embedding space are semantically similar in the real world. Then we build user embeddings by aggregating embeddings of items in the user's recent purchase history. The similarity of user embedding and item embeddings can be calculated to infer user-item affinity and then refine the ranking results in the first stage. After the model got deployed in the production environment, we have observed a significant lift of mean reciprocal rank, which measures the quality of ranking results.

In the next step, we are going to further optimize the personalized ranking model by capturing user's long-term interests and short-term interests. Also, we will keep exploring learn-to-rank models based on the combination of shallow features and deep embeddings.

Snap Analytics

Company: SOTI Inc.
Student: Sara Golestaneh, Ronak Patel, Tony Wei
Industry Supervisors: Amir Mahmoudzadeh, Hossein Taghinejad
Academic Supervisor: Gerald Penn

Mobile apps have become a norm in everyday lives whether it is for productivity at work or for entertainment. Enterprises are finding it seemingly expensive in terms of money and time to hire developers in order to build an app. SOTI Snap is a cross-platform mobile app development solution built for businesses. It empowers organizations to build mobile apps they need with drag and drop widgets and using blocks to create control sequences eliminating the need for development experience. As data gathering and extraction becomes more prominent, so does data analytics and querying. To eliminate the technical barrier required to perform data analytics and querying, Snap Analytics will be an additional feature available within the SNAP platform.

Our project focuses on integrating a question answering system into SOTI Snap which enables non-technical users to ask data related questions and receive the corresponding information from the system. This particular type of question answering is over structured RDBMS which is done by mapping the natural language query utterance to a SQL query representation. To solve this task, our project uses multiple BERT based classification models to predict and construct different portions of the SQL query. Then, the query is executed on the database and the answer is returned to the user by converting the SQL query to NL using a transformer-based model. One important future step of this project will include designing a conversational engine so users can interact with the system asking subsequently interrelated questions.

How do we engage with our Health?

Company: League Inc.
Student: Puneet Gupta
Industry Supervisors: Mehrsa Golestaneh, Oren Kraus
Academic Supervisor: Scott Sanner

Recommendations serve to personalize a platform to your needs and desires. We are told what is good for us, but at the end of the day, our health is personal and we engage with it in a very personal manner. So it begs the question, can we increase our member's engagement with their health through personalized recommendations?

On the League marketplace of health products, we took the purchase history and viewing habits of our members and trained a linear collaborative filtering model. Using a closed form solution to a linear regression problem, we generated an item-item similarity matrix. We determined how much purchasing a service acted as a signal for purchasing another service. Our model was able to identify clusters of services that were similar and signaled a preference.

Using recommendations from the model, we showed that we achieved personalization to our members by beating the Top@K benchmark by more than 150%. The model predicted over 30% of purchases that were held out of training, by serving just five recommendations per user.

Moving forward, we would like to integrate more data sources that provide a personal health context for our members, and see if we can achieve a measurable improvement in health engagement through personalization. Additionally, testing the model in a live, online setting would provide further validation on its utility.

Applied Machine Learning for Early Detection of Retinal Toxicity

Company: Kensington Vision and Research Centre
Student: Faisal Habib
Industry Supervisor: Tom Wright
Academic Supervisors: Arvind Gupta, Huaxiong Huang

Hydroxychloroquine (HCQ) is an anti-inflammatory drug in widespread use for the treatment of systemic auto-immune diseases. Vision loss caused by retinal toxicity is a significant risk due to long-term HCQ therapy. Identifying patients at risk of developing retinal toxicity can help prevent vision loss and improve the quality of life for patients.

Multifocal Electroretinogram (mfERG) is one type of diagnostic test carried out to determine the condition of the eye. However, the interpretation of the results is mostly qualitative and requires an expert analysis. The goal of the research project is to: (i) study and identify the features of the mfERG signals that are most sensitive to retinal toxicity and (ii) develop a machine learning model that can assist medical professionals to identify cases of retinal toxicity more accurately.

We extract mfERG signals from the test instrument and process them to remove noise before extracting features in the time and frequency domains. We use random forests (decision trees) to identify important features and train an ensemble of SVM classifiers. The aggregate result of the classifiers determines the classification of the eyes. Our method results in a sensitive classifier (83%) with an overall accuracy of 88%.

We deploy the ensemble model in a custom application making use of Docker containers.

Intelligent Document De-noising and Ill-conditioned Text Recognition

Company: Vanguard Investments Canada
Student: Yunhao Ji
Industry Supervisor: Ashish Bansal
Academic Supervisor: Florian Shkurti

Vanguard is one of the world's largest investment management companies. As of August 31st, 2020, Vanguard managed USD 6.6 trillion in global assets worldwide. Vanguard Center for Analytics and Insights (CAI) is comprised of several dedicated machine learning teams that not only develop in-house AI-based solutions for specific Vanguard needs, but also conduct cutting edge research on speech, knowledge graph, natural language processing, reinforcement learning, computer vision and more. Business document processing is a key part of Vanguard's daily operation. Robotic Process Automation (RPA) is an intelligent software agent that shifts the human-centred document processing tasks to partial or full automation. The first step of RPA is to convert PDF documents into searchable strings, which is known as Optical Character Recognition (OCR). Open-source OCR engine, such as Google Tesseract, does a good job converting PDF to text when the document is clean but fails in the following scenarios: (1) non-text objects, such as signatures, logos; (2) background noise, such as pepper noise, watermark, coffee stains and highlighters; (3) ill-conditioned text, such as handwriting, text with discontinued strokes (dashed text). In this project, we first de-noise the documents by using (1) Object Detection to identify and subsequently remove, non-text objects, (2) CycleGAN to remove background noise. De-noising improves the Google Tesseract output for our custom document dataset. And for (3), we train a gated-CNN model, achieving character error rate (CER) of 8.01% and 0.5% for handwriting and dashed text respectively.

Building a Multilingual B2B Supplier Domain Classifier

Company: Tealbook
Student: Khasir Hean
Industry Supervisor: Geoffrey Peddle
Academic Supervisor: Gerald Penn

Many of the world's large business-to-business (B2B) suppliers operate English-language website domains, or at least offer an English translation of their web domain. However, a significant portion of B2B suppliers do not offer English versions of their domains, especially small and regional suppliers. Procurement is traditionally a complex and time-consuming process, and this is made even more difficult when seeking foreign-language suppliers. Tealbook, a data-driven procurement platform with over 4 million supplier profiles in its database, wants to speed up this process by enabling buyers to rapidly find and connect with suppliers anywhere in the world.

Our current English-only B2B supplier domain classifier utilizes a combination of various BERT models (Bidirectional Encoder Representations from Transformers) on text extracted from domain homepages for classification. The purpose of this project is to create a classifier which can classify domains in foreign languages such as German, French and Mandarin Chinese, and which offers improvements in performance as compared to the original classifier. Project investigation comprises: 1) the addition of a transformer-based machine translation component to the classifier; 2) the creation of a data pipeline, involving extraction of proprietary as well as public data for model training; and 3) the replacement of BERT models with multilingual-BERT (mBERT) models for improved linguistic generalization ability. Initial results are quite effective thus far. Once research on the multilingual classifier is complete, future research will investigate extracting information from foreign-language supplier domains to further augment tealbook's database of supplier profiles.

Speech Representation for Intent Recognition

Company: Google
Student: Yizhan Jiang
Industry Supervisor: Ye Yuan
Academic Supervisor: Jimmy Ba

Recognizing intent, which categorizes an end-user's intention for one conversation turn, is one of the primary goals of virtual agents. The traditional voice virtual agent uses a two-stage approach to recognize intents: (1) converts speech to text, (2) uses text to recognize intents. However, this approach can have higher latency and more complicated maintenance.

The goal of the project is to explore both supervised and self-supervised approaches to build end-to-end speech-to-intent models. Both approaches aim to create high-quality speech features, which are fed into the same attention-based classification layer for recognizing intents. The supervised approach creates speech features using the encoder from popular automatic speech recognition (ASR) models, such as RNN Transducer (RNN-T) and Listen, Attend and Spell (LAS). In the self-supervised approach, six representative self-supervised speech models are implemented and used to create speech features. The implemented models cover the major types of self-supervised speech learning: Contrastive Predictive Coding (CPC), Autoregressive Predictive Coding (APC), and Masked Predictive Coding (MPC).

We find the supervised approach outperforms the self-supervised approach most of the time unless there is insufficient labeled data in pre-training. In addition, we demonstrate that our best end-to-end speech-to-intent model is comparable to a strong two-stage speech-to-intent model. As future work, since the pre-training data and the intent recognition data can come from different domains, we will explore domain adaptation strategies to improve model generalization to unseen domains.

Learning Latent Representations in Speech for Data Augmentation

Company: Vanguard Investments Canada
Student: Shakti Kumar
Industry Supervisors: Jithin Pradeep, Hussain Zaidi
Academic Supervisor: Gerald Penn

Vanguard is one of the world's largest investment management companies. With a worldwide presence, Vanguard receives millions of calls from its clients across the globe. Transcribing them with minimal error is a challenge due to varying speaking styles and noisy environments in the received phone calls. To improve our in-house speech transcription solution for different English speaking-styles such as British and Australian, we require vast amount of labelled data. Lack of high-quality datasets is one of the main challenges in building a high-performance recognition system. We propose a speech augmentation method for creating such datasets inspired by recent architectures, which synthesize (text, speech) pairs with controllable attributes such as pitch, speaking rate, and pause duration. The existing methods either need many hours of data before they can discover such meaningful features or restrict the number of features discovered, resulting in difficult and limited control on the synthesized speech. To alleviate this problem, we propose an identity map reordering for transformer-based architectures and minimizing mutual information in latent space to better understand the representations of the target demography. Our proposed method discovers compact clusters of speaking styles and synthesizes diverse accents. We surpass the performance of existing LSTM and Transformer based architectures (5% DB Index improvement and 13% Dunn Index improvement on speech data with US and UK accents). We also demonstrate significant stability coupled with accelerated convergence compared to LSTMs while training our model.

Video Spatial Recognition

Company: SOTI Inc.
Students: Qinyu Lei, Rahul Shekhawat, Lipai Xu
Industry Supervisor: Lorne Wilkinson
Academic Supervisors: Anthony Bonner, Sven Dickinson, Kyros Kutulakos

Drones are receiving significant attention as new and fast-developing technology and are being used in various fields, e.g. surveillance, manufacturing, agriculture, construction, etc. Mapping and navigation remain the core problems of drone design. Our project aims to design a comprehensive indoor mapping solution for drones. Our solution would also locate and identify objects along with building 3D maps of indoor environments. The drone is expected to be used for applications, such as search-and-rescue operations for caves & mines, to help the elderly in their homes, home security, and for the creation of high-fidelity floorplans.

Our solution design uses Modularly ADaptive Network (MADNet) for depth perception, Real-Time Appearance Based Mapping (RTAB-Map) for Simultaneous Localisation & Mapping (SLAM), Tiny YOLO for 2D object detection and FrustumPointNet for constructing 3D bounding boxes. We use a ZED2 stereo camera with NVIDIA Jetson Xavier NX board on the drone and a dedicated off-board GPU on the ground station. MADNet can achieve a low error rate (~5%) and improves as it adapts to the environment. RTAB-Map is able to achieve low error in trajectory prediction and produce reliable 3D maps with depth information. Tiny YOLO can achieve near real-time performance on the drone and real-time performance on off-board GPU based ground station. Our next step is to improve the accuracy of the current solution by iterating over different choices of parameters and components, such as feature detectors. The improved solution would be deployed on an actual drone, which is in manufacturing right now, to test it in real-world scenarios. A potential loss in performance is expected outside lab conditions, thus, post-deployment we would be working to fine-tune the parameters for better performance.

Measuring Sentence Semantic Similarity with Latent Categorical Mixtures

Company: RSVP.ai
Student: Minghan Li
Industry Supervisors: Jimmy Lin, Luchen Tan
Academic Supervisor: Animesh Garg

Measuring sentence semantic similarity using pre-trained language models such as BERT generally yields unsatisfactory zero-shot performance, and one main reason is ineffective token aggregation methods such as mean pooling. In this paper, we demonstrate under a Bayesian framework that distance between primitive statistics such as the mean of word embeddings are fundamentally flawed for capturing sentence-level semantic similarity. To remedy this issue, we propose to learn a categorical variational autoencoder (VAE) based on off-the-shelf pre-trained language models. We theoretically prove that measuring the distance between the latent categorical mixtures can better reflect the true sentence semantic similarity. In addition, our Bayesian framework provides explanations for why models finetuned on labeled sentence pairs have better zero-shot performance. We also empirically demonstrate that these finetuned models could be further improved by Latte-Mix. Our method not only yields the state-of-the-art zero-shot performance on semantic similarity datasets such as STS, but also enjoy the benefits of fast training and having small memory footprints.

Automated Parallel Feature Search & Hyperparameter Optimization

Company: Layer 6 AI
Student: Jiahuang Lin
Industry Supervisor: Barum Rho
Academic Supervisor: Gennady Pekhimenko

In recent years, machine learning has been applied in various fields and shown promising results. Researchers have found that finding a proper set of features and tuning a machine learning model can greatly improve the performance of the model for specific AI tasks. However, both feature selection and hyper-parameter tuning can be time-consuming and affected by the subjective view of data scientists. We propose an automated parallel approach for optimizing features and hyper-parameters using statistical methods and simultaneously exploit the large-scale cluster computing power of Kubernetes over Azure to help us shorten the search and optimization time. In practice, we find that this approach can dramatically reduce the number of features while optimizing the performance of the model. In the future, we will adapt our current implementation to make it more generalized to support language agnosticism and better communication between different nodes during runtime. We will also experiment with more statistical methods to explore better algorithms.

Smoke Alarm Detection on the Edge

Company: Ecobee
Student: You-Syuan (Judy) Liou
Industry Supervisor: Inayat Khosla
Academic Supervisor: Roger Grosse

Ecobee is a smart home company that brought the first smart thermostat to market in 2009. Since then, it not only put effort into providing users more comfortable homes but also safer ones. Recently, it has started bringing home monitoring service to detect events happening at home and send out instant notifications for unexpected ones, like unusual entries and sounds.

This project aims to enhance home monitoring service with an acoustic feature, smoke alarm detection, that runs on the edge. Once the smoke alarm sound is detected at home, the user will receive an alert on his mobile. Given the sounds of smoke alarms varies between different brands, our work is to distinguish the sound based on its pattern, three beeps followed by a one-second silence. We convert audio to spectrograms and develop an audio classification model with Convolutional Neural Networks (CNN). To make it runnable on the devices with much less computation power, we leverage Tensorflow Lite, a dedicated library for ML on the edge, and build an end-to-end flow with audio filter as the first defense layer to minimize the times to call the CNN model.

This approach achieves a F1-score of 88.9% and successfully runs on our devices. Also, during this project, we build an audio data augmentation pipeline to fasten acoustic feature development. Our next step is to build multiple acoustic detection features within the hardware constraints.

Thread Representation for Online Forum Recommendation

Company: VerticalScope Inc.
Student: Ding Tao (Danny) Liu
Industry Supervisor: Phileas Hocquard
Academic Supervisor: Gerald Penn

Online forums with millions of threads present an information overload to users who are querying for specific discussions. The performance of recommender systems and search engines is crucial for a positive user experience. This project seeks to learn compact thread representations that power a content-based recommender system for discussion forums. A content-based recommender system finds content that is similar to what the user likes, thus the thread representations must capture semantic textual similarities. Threads also exist across many different topic domains, so the representations should remain relevant when compared between domains as well as within domains. There are existing text embedding models that function at the word, sentence, and document levels. However, the online thread medium presents some unique challenges. Context drift is the problem where the topics, or context of discussion deviates from the main post as the sequence of replies grows. With a large user base, out of vocabulary words are also very common. Lastly, threads can vary drastically in length, from tens of posts to thousands of posts, and so having a fixed size representation may lose information. Due to different granularities of information, threads can be broken down into the title, the main post, and the body. Each of these components are represented separately, and a distance function that takes all three representations is used to compute similarity. A hierarchy of embedding models can be created by training on data from different domains to support inter- and intra-domain representation comparisons.

On the Convergence of Continuous Constrained Optimization for Structure Learning

Company: Mila
Student: Ignavier Ng
Industry Supervisor: Yoshua Bengio
Academic Supervisor: David Duvenaud

Structure learning of Directed Acyclic Graphs (DAGs) is a fundamental problem in many scientific endeavors. A new line of work, based on NOTEARS (Zheng et al., 2018), reformulates the structure learning problem as a continuous optimization one by leveraging an algebraic characterization of DAG constraint. The constrained problem is typically solved using the Augmented Lagrangian Method (ALM) which is often preferred to the quadratic penalty method (QPM) by virtue of its convergence result that does not require the penalty coefficient to go to infinity, hence avoiding ill-conditioning. In this work, we review the standard convergence result of the ALM and show that the required conditions are not satisfied in the recent continuous constrained formulation for learning DAGs. We demonstrate empirically that its behavior is akin to that of the QPM which is prone to ill-conditioning, thus motivating the use of second-order method in this setting. We also establish the convergence guarantee of QPM to a DAG solution, under mild conditions, based on a property of the DAG constraint term.

Weakly Supervised Named Entity Recognition

Company: Vanguard Investments Canada
Student: Jerrod Parker
Industry Supervisor: Shi Yu
Academic Supervisor: Bo Wang

Weakly supervised methods estimate the labels for a dataset using the predictions of several noisy supervision sources. Many machine learning practitioners have begun using weak supervision to more quickly and cheaply annotate data compared to traditional manual labeling. This project focuses on the specific problem of weak supervision for Named Entity Recognition (NER). Current methods for weakly supervised NER either do not make use of context to resolve disagreements, discard useful information about the weak labels such as the number of votes per class and identities of the voters, or use a separate algorithm to convert the problem to i.i.d classification. To improve upon each of these shortcomings, we propose an end-to-end model that learns the contexts where each weak labeler is the most accurate and uses this knowledge to estimate the true labels. Results of experiments on several weakly supervised named entity recognition datasets show substantial improvements in F1 score over the state-of-the-art when there are many disagreements between the weak labelers.

End-to-end Machine Learning Pipeline for Harsh Accident Detection

Company: Geotab
Student: Amish Patel
Industry Supervisosr: Willem Petersen
Academic Supervisor: Marsha Chechik

Road safety affects everyone, not just Geotab customers. In Canada alone, according to the Canadian Automotive Association, the economic cost associated with road crashes is approximately $25 billion per year. Geotab is a global leader in the telematics industry which has been collecting data from over 2 million connected devices over the past few years. As such, the objective is to use the telemetry data to classify accidents through semi-supervised learning. The reason behind semi-supervised learning is because the labeled data is extremely small as opposed to the millions of potential events. The solution to this research problem is using a multi-step approach by building a complete end-to-end pipeline which consists of: (1) Gathering all interesting events by using a threshold of 2.5G linear acceleration magnitude (2) Filtering out all false positives using domain knowledge and general insights found during exploratory analysis which includes bad installs, non-stationary events, door slams, and potholes (3) Engineer features from acceleration, GPS, and engine data that help in classifying accidents (4) Heuristic-based accident classification (5) Semi-supervised based approach. The outcome of the heuristic based approach results in 0.73 AUC. The outcome of the initial autoencoder and gradient boosting trees model results in mean AUC of 0.85 using 10-fold cross validation. In future, the plan is to use deep learning techniques (Time-Aware LSTM and Neural ODE) to get closer to a real-time accident classification using only a localized two second window around the event time.

Semantic Clustering Search Engine for Suppliers

Company: Tealbook
Student: Nayantara Prem
Industry Supervisor: Geoffrey Peddle
Academic Supervisor: Gerald Penn

Tealbook helps organizations discover suppliers with its curated database of around 4 million suppliers and a search engine that presents hundreds of relevant suppliers to a search query. The current search engine finds relevant documents through the traditional keyword search approach based on the occurrence of the query terms in the documents being searched and displays the results in a list to the user. This approach often falls short of capturing the user's intent because of language ambiguity in the query and documents. The large number of results exacerbates the user's ability to find actually relevant suppliers.

The goal of this research is to make it easier for users to navigate this large set of search results to find the semantically relevant suppliers they need. We propose a search clustering engine that organizes the results of a user search into groups of suppliers providing semantically similar services and products, with each group described by a meaningful label. In order to measure supplier similarity from the text on supplier websites, we learn a set of unsupervised representations of each supplier by clustering sentence embeddings created using Google's deep neural model, BERT. We apply a modified mutual information algorithm to label the final clusters to clarify their content to the searcher. In the future, we can use this multi-facetted supplier representation for other stages in supplier chain management outside of supplier discovery.

iStandardize: Recommendations for Healthcare Form Standardization

Company: Deloitte
Student: Joe Roussy
Industry Supervisor: Lynn Luo
Academic Supervisor: Michael Brudno

iStandardize is a Natural Language Processing (NLP) solution that is designed to streamline the standardization of clinical order sets (i.e., forms). Currently, hospital networks use multiple versions of forms and order sets, many of them similar in nature. The lack of standardization poses challenges in integrating the data for sharing, adds additional documentation burden, and disrupts the workflow for clinicians. The solution applies current NLP and Machine Learning (ML) techniques to identify similar order sets and their elements (attributes and responses), reduce the manual work required to compare the order sets, and expedite the decision-making process for standardization.

A PoC version of the solution was enhanced by improving the quality of the clusters created and a visualization was developed to highlight the insights of the generated recommendations. In the PoC, clusters were created by applying hierarchical clustering to cosine similarity matrices of word embeddings. Cluster quality was improved by incorporating custom non-semantic contextual features into the similarity calculations. In this enhanced version, the similarity matrix used for clustering is a weighted average of the word embedding and contextual feature similarity matrices.

Over the course of the internship, iStandardize was used by a client organization in a standardization project resulting in an estimated cost savings of $750k – $1M. Next steps include further improving the quality of recommended clusters by developing additional contextual features and weighting medical terms more in similarity calculations. The solution will also be developed into an automated end-to-end processing pipeline.

Disentangling Latent Space for Hair Style Transfer

Company: ModiFace Inc.
Student: Rohit Saha
Industry Supervisor: Brendan Duke
Academic Supervisor: Florian Shkurti

Hairstyle transfer is challenging due to hair structure differences in the source and target hair. Therefore, we propose Latent Optimization of Hairstyles via Orthogonalization (LOHO), an optimization-based approach using GAN inversion to infill missing hair structure details in latent space during hairstyle transfer. Our approach decomposes hair into three attributes: perceptual structure, appearance, and style, and includes tailored losses to model each of these attributes independently. Furthermore, we propose two-stage optimization and gradient orthogonalization to enable disentangled latent space optimization of our hair attributes. Using LOHO for latent space manipulation, users can synthesize novel photorealistic images by manipulating hair attributes either individually or jointly, transferring the desired attributes from reference hairstyles. LOHO achieves a superior FID compared with the current state-of-the-art (SOTA) for hairstyle transfer. Additionally, LOHO preserves the subject's identity comparably well according to PSNR and SSIM when compared to SOTA image embedding pipelines. Code is available at https://github.com/dukebw/LOHO.

SOTI Snap Widget Enhancements

Company: SOTI Inc.
Student: Weijia Sun
Industry Supervisor: Shash Anand
Academic Supervisor: Fanny Chevalier

Companies nowadays are using mobile apps to digitalizing their paper work process, which requires time and labor cost for development and maintenance of such apps. SOTI Snap provides a no-code/ low-code development platform for anyone without prior programming knowledge to build apps using drag-and-drop within few hours. Specifically, a current widget for displaying records in a list can enable users to sort, edit and delete records in this table. However, if users have thousands of records, it may be problematic to display all of them in such a small screen on mobile devices. Also, it is difficult to find certain records in this table. Design thinking is adopted here as the methodology to provide scientific framework and guidance for understanding users' pain points and generating a better solution. Furthermore, technical hurdles are addressed in the implementation stage, which involves Android agent, server, front end and back end. In the next version of SOTI Snap, this enhanced widget will have pagination function for users to change records per page in a single screen, together with the choice of style and position. Besides, search function is added to enable user's ability of searching across all records. In the future, A/B test will be used to find the next elements to be enhanced using statistics data collected from users.

Unsupervised Super-Resolution Training Data Construction

Company: Samsung AI Centre
Student: Xinyu Sun
Industry Supervisor: Alex Levinshtein
Academic Supervisor: Animesh Garg

Super-Resolution (SR) aims at predicting image details of the given low-resolution (LR) images to retrieve the corresponding high-resolution (HR) counterparts. Although many super-resolution methods have demonstrated good performance on experimental datasets, they often fail in real-world image super-resolution tasks. One of the reasons is the widely used degradation method (i.e. bicubic downsampling) for constructing Low-Resolution (LR) and High-Resolution (HR) training pairs cannot capture the complicated image degradation process in the real world. Therefore, constructing real paired data for Super-Resolution becomes a key problem. Other methods address this issue by using a more realistic synthetic downsampling or bridging the gap between synthetic and real LR images using GANs. Instead, we use an unsupervised domain adaptation framework to construct synthetic HR images for pseudo ground truth. We first upsample LR images using a state-of-the-art blind super-resolution method. Then we use a CycleGAN-based model to transfer the upsampled images to the real-world image domain and use them as our ground truth to train our super-resolution model. Our experimental result shows that our final model achieves sharper results than the standard model which trained on bicubic downsampled LR images, and yields fewer artifacts compared to methods that use synthetic LR images for training. In the remaining term of research, we will focus on the optimization of our model.

Intelligent Code Completion for SOTI Blockly

Company: SOTI Inc.
Students: Xuejie (Alicia) Tang, Feilong Song
Industry Supervisor: Shash Anand
Academic Supervisors: Maryam Mehri Dehnavi, Eyal de Lara

SOTI Blockly is a block-based visual programming environment supported by SOTI SNAP. It helps non-technical users to add programming logic by simply dragging and dropping logic blocks onto a canvas, thus building powerful, visually compelling apps in just minutes. Intelligent code completion is a context-aware code completion feature in some programming environments that speeds up the process of coding applications by reducing typos and other common mistakes. Attempts to do this are usually done through auto completion popups when typing, querying parameters of functions, query hints related to syntax errors, etc..

Here, our goal is to introduce the code completion concept into the visual programming environment SOTI Blockly using machine learning methods. We provide block prediction functionality, which mimics code autocompletion, to predict the next possible blocks according to the current contexts. First, we introduced a data pipeline including collection service, model training service, and prediction services with Apache Spark Streaming. An Abstract Syntax Tree (AST) based transformer model was used for predicting the blocks. We evaluated the model on a dataset of Scratch programs and reached an accuracy of 90.25% of targeting blocks appearing in the top 5 results. Future work includes further improving the accuracy, reducing the inference resources, supporting the prediction of complicated block combinations, and improving the pipeline's efficiency of transmitting and processing data by optimizing the state transferring model in Spark Streaming context.

Deep Learning for Fundamental Forecasting

Company: CPP Investments
Student: Xuxu Tang
Industry Supervisor: Yixing Zhang
Academic Supervisor: Radu Craiu

Machine learning models have become widely used in investment strategies to generate investment returns. As stock prices fluctuate due to various factors, stock returns are noisy and hard to predict. Studies have shown that incorporating fundamental forecasts in investment strategies can help generate long-term investment returns. Since publicly traded companies disclose their financial information periodically, the financial features should have time-series characteristics. As a result, the objective of this project is to explore deep learning models on fundamental forecasting to capture nonlinear time-series information. A recurrent neural network can extract time-series relationships in a dataset. However, due to the vanishing gradient problem, RNN cannot capture long-term dependencies. To solve this problem, we have explored the LSTM model that captures long-term dependencies by utilizing the cell state, the forget gate, the input gate, and the output gate. After selecting the appropriate features, constructing and tuning the model, we train the model with the training data and forecast the target on the test data. Model performance is calculated by the correlation between forecasted and real target. The output analysis shows that LSTM outperforms linear models such as Lasso. For different time periods, it outperforms non-linear models such as LightGBM. The next steps of the project are to test LSTM models on different types of datasets, obtain model's correlation with other models and construct ensemble forecasts, select the most important features based on the LSTM model, and summarize findings and construct project report.

Biological Spectroscopic Signal Denoising

Company: Synex Medical
Student: Abenezer Teklemariam
Industry Supervisor: Ben Nashman
Academic Supervisor: Adam Stinchcombe

Nuclear Magnetic Resonance (NMR) is a powerful technique used to identify the composition and concentration of individual compounds in a chemical mixture. One such application is the non-invasive monitoring of metabolites in the human body. However, NMR signals are inherently very weak due to the small nuclear magnetization and noise could overwhelm the signal if concentrations are low. For such NMR devices, the noise is composed of thermal and electrical noise as well as Radio Frequency Interference (RFI). At Synex, we developed a Machine Learning algorithm and a state-of-the-art signal acquisition system to recover high fidelity signal from a noisy input. First, we map the signal onto a high dimensional subspace, such that some of the noise is isolated from the signal. Inspired by image denoising techniques, we then use a deep Convolutional Neural Network (CNN) to denoise the high dimensional signal and map the denoised signal back to a 1D subspace. By employing adaptive noise estimation and dilated convolutions in the CNN, we reduced our error by 25%. We plan to incorporate additional information into our model to help us improve our denoising performance and achieve quantification accuracies that surpass industry standards.

Integrating Consul Service Mesh with Kafka within and across Datacenters

Company: Ethoca Technologies, Inc.
Student: Yao Tu
Industry Supervisors: Hrant Arakelyan, Dale Woolridge
Academic Supervisor: Ashvin Goel

Service mesh has been around for a while as a solution for service-to-service communication between microservices. Yet its adoption is most usually seen in request-response style synchronous communication like web application. In this project we want to explore the potential of using service mesh for asynchronous event-driven messaging architecture. We chose Consul as the service mesh implementation and integrated it with a typical event-driven messaging architecture with Kafka. To understand how service mesh works within and across clusters, we implemented the integration in both scenarios. In the multi-datacenter scenario, we took a step further by scaling the VM-based Kafka cluster into Kubernetes. We measured the message throughput before and after the Consul integration to understand the performance implementation of integrating service mesh. We also compared the security model of Consul and Kafka's existing one.

Spoken Language Identification for Children

Company: Pearson Canada Inc.
Student: Aravind Varier
Industry Supervisor: Richard Wang
Academic Supervisor: Gerald Penn

Several speech processing systems use Language Identification (LID) as an initial step in a chain of different components. Traditionally, LID systems were designed by hand-crafting features from audio samples. In the past few years, advances in the fields of Deep Learning have brought about state-of-the-art performance in LID without putting in any effort to hand-craft domain specific features. In this project, Pearson aims to classify languages spoken by children over a short duration of 5-90 seconds into popular languages like English, Spanish and Mandarin. We analyzed the performance of various state-of-the-art systems for the task of spoken language identification, and also experimented with the two most popular deep learning solutions: the X-Vector system and the Convolutional Recurrent Neural Network (CRNN). We first started with techniques to deal with a highly imbalanced dataset. This included experimenting with weighted loss functions (techniques such as class-balanced loss) and changing the sampling ratio. After doing this exploration of the data, we then moved onto modeling. We found that, while the CRNN system dominates in terms of pure performance, the X-Vector system is both faster and lighter and not as sensitive to hyper-parameter values as compared to the CRNN system. We also experimented with a novel Convolutional Transformer architecture for LID that takes lesser time to train for long length sequences and achieves comparative performances. Through this project, we will share our exploration and our experience of applying deep learning as a solution to LID.

Designing 'Zero Credit Touch' (ZCT) pre-approved Credit Underwriting Program for Retail Customers

Company: ICICI Bank
Student: Sanghamesh S. Vastrad
Industry Supervisor: Sandipan Ray
Academic Supervisor: Sebastian Jaimungal

ICICI Bank has developed various 'Zero Credit Touch' (ZCT) strategies where without any credit intervention and additional information taken from customers, credit facilities can be provided. But there are several challenges in the expansion of ZCT strategies, namely, (i) current credit models which are a combination of business rules, scorecards and machine learning models, do not qualify a significant proportion of existing ICICI Bank customers; (ii) wherever customers do not have a salary account with the Bank, estimated income is lower leading to the customer being offered an amount lower than his/her requirement; (iii) customers with fraudulent intentions can open accounts and over time, these profiles would qualify for ZCT.

We develop a novel ZCT system incorporating several state-of-the-art methodologies to tackle each of the problems mentioned above and build one go-to product to reduce credit and operations cost of lending whilst providing a superior customer experience. Most notable methodologies include (i) Synthetic data generation using a combination of SMOTE and Generative Adversarial Networks to address the class imbalance problem in credit underwriting and fraud detection, (ii) Bayesian Neural Network for income estimation to create a confidence interval, (iii) Latent Dirichlet Allocation for customer segmentation based on information availability, (iv) Stacked Ensemble of multiple models for improved prediction. Experimental results show an improvement in Gini Coefficient by over 6% compared to the previously deployed credit underwriting model.

SOTI SNAP Widget Enhancement

Company: SOTI Inc.
Student: Yuhan Wang
Industry Supervisor: Shash Anand
Academic Supervisor: Yashar Ganjali

SOTI SNAP is a cross-platform software that allows anyone to create a mobile app with no programming or technical knowledge. It allows users to create apps by dragging and dropping widgets onto a canvas and connecting them together. The purpose of the project is to enhance existing widgets, brainstorm and build new features so that anyone can build apps easily and effectively.

In SOTI SNAP, an app consists of pages and a page consists of a bunch of widgets. In today's SNAP, customers can only copy and paste a widget on the same page. If they wish to reuse the same customized widgets in a different page or app, they have to build a new widget from scratch. Understanding this limitation brings customer repetitive works, I designed a new feature called Visual Clipboard: a visual place for customers to save and reuse their widgets and pages by easily dragging and dropping.

Besides designing and building visual clipboard, I also worked on enhancing the layout widget in SNAP. The new layout widget in SNAP will give customers more possibilities in organizing widgets. With the new layout widget, the app not only works well, but looks great too.
The next step of the project is working on the IDE side of the layout widget. Building a theme gallery for customers to choose layout themes when they wish to explore creative ways of organizing widgets.

Machine Learning for Cancer Therapy Benchmarking

Company: bridge7 Oncology
Student: David Hong Yang
Industry Supervisor: Chris McIntosh
Academic Supervisor: Marzyeh Ghassemi

Cancer treatment planning may differ between medical institutions due to a lack of imposed standards and inequality in resources. Bridge7 Oncology provides an AI generated benchmark score for cancer therapy plans. The model uses a gold standard dataset as reference and analyzes 3D medical images along with data quantifying the patient's treatment to determine whether the prescribed treatment is a good match for the patient. One major limitation with the current model is that the lack of gold standard data for uncommon cases such as heart or kidney cancer leads to poor model generalization for these cancer types. The goal of this project is to develop a model that can still generalize well with a scarcity of data for certain cancer types. To address this problem, we propose to train a model utilizing meta-learning, treating each cancer type as a task. We evaluate the models trained using two meta-learning algorithms, MAML and Reptile, on clinical data. In addition, other feature embedding and data representation techniques will also be explored to improve overall model performance and generalization.

Natural Language-Based Live Video Filtering Proxy

Company: Samsung AI Centre
Student: Ning Ye
Industry Supervisor: Iqbal Mohomed
Academic Supervisor: Eyal de Lara

Rapid breakthroughs in deep learning have shown promising results in intelligent video analytics in the cloud. However, as cameras at the edge may be bandwidth-constrained, it is not viable to upload all the video frames to the cloud for further processing. This research project focuses on matching live video streams to a set of predefined queries specified by natural language to effectively filter out irrelevant frames. This can serve as a powerful tool for a multitude of applications. Home robots could be programmed to notify the user when specific events are triggered, such as a dog scratching the couch, or a child playing with the stove. Existing systems that support video-text matching typically map the video and text data into a joint embedding space to compute the similarity scores. However, the video features are extracted offline and evaluation is done over the entire video. This makes it infeasible for the model to work in real time while only analyzing a subset of incoming frames. Building on top of a state-of-the-art video-retrieval framework, we have proposed a lightweight architecture to filter incoming frames based on user queries of interest. Not only can the system generalize to complex, unseen queries without retraining, it also improves the event detection accuracy of the state-of-the-art filtering approach by 1.5 times. As a next step, we will explore techniques to increase the number of feature extractors run in real time.

Building Mobile Apps Using SOTI Blockly

Company: SOTI Inc.
Students: Raymond Zeng, Mingqi Zhu
Industry Supervisor: Shash Anand
Academic Supervisor: Fanny Chevalier

SOTI Blockly is a block-based programming environment built off of Google Blockly. It uses visual programming to allow users to drag and drop blocks to create the logic they need. SOTI Blockly has been integrated into SOTI SNAP, a mobile application development tool. SOTI Blockly supports programming functions such as math operations, text manipulation, and loops to help design mobile application logic.

Previously, logic created in SOTI Blockly could not be saved and reused in different mobile apps. We researched other block-based programming environments, such as Scratch and Microsoft MakeCode, and conducted interface surveys within our group (N=16). Using our research, we developed a React-based dashboard and used Node.js and PostgreSQL to construct a three-step block creation process. This process allows users to customize their own blocks and reuse logic elsewhere.

Custom blocks can be created using a combination of existing blocks, or by defining the behaviour of the block using JavaScript. This makes SOTI Blockly much more versatile and lets the user create a greater variety of apps. We also developed "My Libraries," a block management tool allowing users to organize their custom blocks as they create more and more blocks. Users can share their libraries, helping to maintain the visual programming aspect of Blockly, as future users will not have to recreate the logic for those blocks again. Further usability studies (N=16) were conducted to determine an optimal design for a marketplace where users can simply import libraries of blocks for the logic they need.

Indoor Geofence

Company: SOTI Inc.
Student: Zichun Zhuang
Industry Supervisor: Dmitry Shesterin
Academic Supervisor: Eyal de Lara

Many mobile devices that are used in an enterprise are considered mission critical as they are heavily relied upon, allowing a worker to do their job. Locating these devices anywhere and anytime is important. SOTI's technology leverages GPS on a device to track its location and can also alert if a device leaves/enters a predefined geographical fence on a map, otherwise known as geofencing. Outdoor and GPS based on geofencing is widely adopted by many SOTI customers, yet indoor location and geofencing remain relatively novel. This project is interested in researching the indoor location and geofencing capabilities based already deployed enterprise Wi-Fi. This approach relies on existing infrastructure and does not require installation of additional sensory hardware and avoids the corresponding maintenance overhead.

Our work is mainly focused on:
- Reducing the variance and outliers of the current indoor positioning system
- Improving the accuracy of positioning
- Minimizing the number of false alarms of geofences
- Developing a scalable geofence algorithm.
- Optimizing a Point in Polygon algorithm to support large scale geofence

Due to the difficulty of getting the ground truth location of moving devices, we also propose a test environment, a robot that leverages lidar to get its location in real-time. This not only simplifies the testing but also provides us with ground truth to have a standard metric for our experiments.

Dynamic Entity Verification for Trade Documents

Company: Scotiabank
Student: Pengyu Chen
Industry Supervisor: Pragnya Addala
Academic Supervisor: Gerald Penn

The Confirmation Team at Scotiabank is responsible for the verification of contracts for trades such as equity swaps, interest rates swaps, etc. The process involves 3 steps:

1. Two parties determine a trade to be made and a template for constituent terms.
2. One party creates an agreement
3. The second party reviews the agreement to ensure it adheres to the template.

There is significant possibility for changes in both numbers and textual terms, creating great risk for the bank. This risk is addresses by employing human agents and the combination of an AI solution, namely AIDOX, provided by Digital Banking and Operations Analytics.

Previously, AIDOX treats a document as plain text, blindly compares it against the template and then output textual discrepancies to agents for further review. However, certain entities are naturally dynamic and should not be verified as plain text. For instance, contracts drafted on different dates have different termination dates, and therefore are not expected to match those in templates.

The goal of this project is to enhance AIDOX so that it can handle dynamic entities properly. The difficulty comes from two aspects. Firstly, the dynamic property of an entity is highly dependent on contexts and business knowledge. Second, the system needs to ensure a 100% precision considering the volume of funds involved. With that in mind, we developed a dynamic entity verification system powered by syntactic parsing, sentence similarity and pattern matching, which enable us to impose business knowledge and retain a perfect precision score.

Material and Lighting Estimation Based on 3D Implicit Representation

Company: NVIDIA
Student: Jiehan Wang
Industry Supervisors: Jean-Francois Lafleche, Gavriel State
Academic Supervisor: Animesh Garg

Novel view synthesis has been a popular problem in 3D Deep learnings recently. NeRF is one of the papers that shows the amazing result. It takes multiple images from different views as training data and trains a neural network that given spatial location (x, y, z) and viewing direction (θ, φ) as input, the network outputs RGB values and volume density. Then these values are passed into a volumetric renderer, which can render images from any novel views. However, NeRF is only able to produce RGBs for each spatial location. We can extend the NeRF so that it could produce materials and lighting information that can be used in the standard rendering pipeline.

We could then utilize our Kaolin (3D Deep Learning Library) to get the geometry out and NVIDIA Ominiverse Kit to apply materials and lighting on it.

We modify NeRF's network structure so that it outputs diffuse color, roughness, and metallic for each location. We notice that the most important aspect that affects the final rendering qualities is normals. We adopt the derivative of the volume density as our normal. During the training phase, we take the values output by the network and pass them into use a physically-based BRDF proposed in the Epic Game's paper[1] to do the physical rendering. As for the lighting model, we use 9 coefficients of the spherical harmonics lighting. We use MSE loss between the rendered images and the ground-truth and add a normal regularization loss to ensure the local smoothness of the normals. After fully trained the model, we could use CUDA accelerated Marching Cube implemented in Kaolin to get the mesh of the geometry and visualize it in NVIDIA's Omniverse Kit with all the material and lighting information.

We could produce similar results original NeRF's paper with all the additional materials and lighting information that can be used in the standard rendering pipeline. Please see my slide for visual results/outcomes.

What's Next: 1. Extract materials information as a texture map; 2. Apply materials and lighting estimation to other NVIDIA's research (currently applying to Def-Tet in progress); 3. Add specular lighting on top of the spherical harmonics lighting model.

Automated Diagnostic Solution for Managing Mobile Devices Operations

Company: SOTI Inc.
Student: Zichun Zhuang
Industry Supervisor: Hossein Taghinejad
Academic Supervisor: Murat A Erdogdu

SOTI Insight is a business intelligence solution that collects various device specifications over time, such as battery level, signal strength, temperature and provides analytics on applications, data, network and location usage, and operations. This ensures ease for companies to gain insights into the performance of their mobility deployments. As the complexity of device networks increases, the operation and maintenance tasks become complicated, and as a result, the device is more prone to failures. This lays down the requirement for having self-healing systems i.e. troubleshooting in an automatic way.

The research project would concentrate on identifying the factors that are responsible for poor device performance as well as to detect potential patterns that lead to failures. According to our analysis, five key factors cause device failures and limit performance; weak signal strength, high data usage, high battery temperature, prolonged application-in-foreground, and the derived factor from all these the battery discharge rate. We propose a time series forecasting model that could predict the discharge rate to take a proactive step and alert beforehand, hence avoiding any device failure.

The project can be divided into 4 phases: (1) Exploratory data analysis, clustering and gradient boosting trees with SHAP (for explainability) to extract the Key Performance Indicators (KPI) that can accurately measure and predict discharge rate (2)Finding the quantitative relationship between each of the KPIs and discharge rate using data smoothing and statistical modelling (3) Exploring machine learning models for time series forecasting and statistical analysis to accurately model each of the factors and give a formula for discharge rate. (4) Validating the formulas generated in the second phase by running controlled experiments on various devices. This will help the customers identify any of the problematic factors and alert beforehand to avoid any device or network failures.

User-Controllable Selective Paraphrase Generation

Company: RSVP.ai
Student: Mohan Zhang
Industry Supervisor: Luchen Tan
Academic Supervisor: Animesh Garg

The problem that our model aims to solve is paraphrase generation. As the name suggests, we want to automatically generate paraphrases for a given sentence, expressing the same meaning with different syntax. In the paraphrase generation task, source sentences often contain phrases that should not be altered. These phrases which cannot be altered, however, can be context dependent and can vary by application. Our solution to this challenge is to provide the user with explicit tags that can be placed around any arbitrary segment of text to mean "don't change me!" when generating a paraphrase; the model learns to explicitly copy these phrases to the output. The contribution of this work is a novel distant supervision and data generation technique that allows us to start with a standard sequence-to-sequence model and fine tune a paraphrase generator that exhibits this behavior, allowing user controllable paraphrase generation. Additionally, we modify the loss during fine-tuning to explicitly encourage diversity in model output. Our technique is language agnostic, and we experimented on both English and Chinese.

C-Learning: Horizon-Aware Cumulative Accessibility Estimation

Company: Layer 6 AI
Student: Panteha Naderian
Industry Supervisors: Harry Braviner, Gabriel Loaiza-Ganem
Academic Supervisors: Animesh Garg, Chris Maddison

Multi-goal reaching is an important problem in reinforcement learning. It is a suitable framework for advanced real-world agents that need to interact with a diverse set of tasks. Despite recent advances in this field, current algorithms suffer from three major challenges: 1) high sample complexity due to the need for a large number of interactions with the environment; 2) learning only a single way of reaching the goals; and 3) difficulties in solving complex motion planning task due to the the problem of finding the shortest path to a goal in real-time. In order to address these limitations, we introduce the concept of cumulative accessibility functions, C-functions, which measure the reachability of a goal from a given state within a specified horizon. We show that these functions obey a recurrence relation, which enables learning from offline interactions. To learn C-function, we proposed a training method called C-learning. We evaluate our approach on a set of multi-goal discrete and continuous control tasks. We show that our method outperforms state-of-the-art goal-reaching algorithms in success rate, sample complexity, and path optimality. Additionally, our method can trade-off speed and reliability in goal-reaching by suggesting multiple paths to a single goal depending on the provided horizon.

Enhanced Content-Based Similarity Detection for Book Recommendation

Company: Rakuten Kobo Inc.
Student: Peilin Sun
Industry Supervisor: Chris Sjostrom
Academic Supervisor: Murat A. Erdogdu

At Kobo, item recommendation is the main way users discover books on its e-book and audiobook platform. It is based on users' purchase history and reviews, through a process known as collaborative filtering, which provides a good approximation for "similarity". As this does not give ideal results for books with little purchase history, in this project, an advanced book recommender system has been developed based on Kobo's existing content-based similarity detection algorithm.

With book descriptions as the training data, two approaches have been implemented, Latent Dirichlet Allocation (LDA) and doc2vec, to extract the embedding vectors of books. Recommendations were generated by comparing the similarity scores between all book pairs. Considering the possibilities of production, as well as low time complexity and high parallelization potential, the doc2vec model was chosen as the final model. By comparing the recommendation results with the original model and purchase-based similarities, the new approach worked significantly better than the original one, with doubled mean average precision scores. For further improvement, future work includes considering name entity recognition in text preprocessing and leveraging more book content. Finally, an A/B test was performed on production, with statistically significant result in terms of click conversion rate.

Multi-Person 3D Pose Estimation from Multiple Views

Company: Surgical Safety Technologies
Student: Priya Thakur
Industry Supervisor: Frank Rudzicz
Academic Supervisor: Animesh Garg

The task of determining the position and orientation of a person's body parts in an image has great potential in surgical activity recognition. Hence, over the past few decades, 3D human pose estimation (HPE) has received significant attention from the computer vision community and the scientific community. However, 3D pose estimation of multiple humans from multiple views remains an unsolved problem because of huge state space, occlusions and difficulty at establishing cross-view correspondences among noisy and incomplete 2D human pose predictions. This research seeks to address the issue of 3D pose estimation of multiple humans from multiple views. The problem is addressed in three steps. First, we use an off-the-shelf 2D human pose detector to detect bounding-boxes and 2D pose coordinates of each person from all the views. Then, in the second step, the poses that correspond to the same person in different views are grouped into clusters according to appearance and geometry cues. The final step is to estimate a 3D pose for each person (i.e. each cluster) by 3D pictorial structure models. Two models, Detectron2 [1] and Cascaded Pyramid Network (CPN) [2] were trained on the MS COCO dataset to detect bounding-boxes and 2D pose coordinates. We test the model on two public datasets, Campus and Shelf. Our 3D HPE model achieves significant results with CPN as the backbone (96.3% and 96.9% on the Campus [3] and Shelf [3] datasets, respectively) with time efficiency improvements, thus making the model possible to run in real-time surgeries.

SOTI SNAP Digital Signage and Video Player

Company: SOTI Inc.
Student: Sijie Han
Industry Supervisor: Shash Anand
Academic Supervisor: Yashar Ganjali

SOTI customers are using paper and pen based processes in many areas of their business because they perceive building mobile apps to be expensive, time consuming and resource heavy. SOTI SNAP is designed to allow anyone to build apps in minutes.

Incorporating videos into mobile applications is a highly common use case that SOTI Snap was missing. This limited the types of apps that can be built, many of which were replacing paper based forms.

My work was to research the various video capabilities within mobile apps, the different use cases, understand how they are designed and then design and develop a simple interface that allows ANYONE to build and incorporate videos in their mobile apps in minutes.

Users can create and configure widgets on the IDE Console which is developed by .NET framework and AngularJS. After publishing the app, all widgets, configurations and libraries will be saved as contract packages and sent to the server. When users login SNAP on their mobile devices, the app will receive the contract packages from server and initialize widgets.

We provide three default player modes for different use cases: fullscreen, window and floating window. Fullscreen and floating window mode are based on device native player while window mode is based on videojs player. Users can switch between each of the modes. The video source can be online or pre-installed on the devices by using MobiControl technologies. Also, we provide a lot of video controls for users: loop videos, allow/disable pause/fast forward/rewind/exit/autoplay and set aspect ratios.

The following work will be:
1. Combine with Blockly in order to allow the player to interact with other widgets.
2. Add another video source: S3 bucket.
3. Provide more video controls/events. (e.g. timing, video end/pause event)

Synthetic Data Generation from Knowledge Graph for Language Modeling

Company: Amazon
Student: Yu-Siang Wang
Industry Supervisors: Volker Leutnant, Prabhat Pandey
Academic Supervisor: Animesh Garg

Understanding and knowing about rare and trending entities plays an important role for a virtual assistant to be perceived as smart. Recognizing such entities is a challenge for the Automatic Speech Recognition (ASR) system due to lack of sufficient text data for language model (LM) training. To tackle the aforementioned problem, we propose a synthetic data generation method using Knowledge Graph (KG). We identify entities in our internal and external text data sources using an in-house semantic parser to derive contexts and exploit the relations between entities and their classes in the KG to generate synthetic texts for the rare and trending entities. We use this synthetic data to augment training sources for both first-pass n-gram LM of the hybrid ASR system and Neural Language Model (NLM) which is used to rescore n-best hypotheses generated from the first-pass system. We evaluate the performance in terms of Word Error Rate Reduction (WERR) over the baseline system on two internal test sets: the first consists of utterances with heavy entity references, and the second focuses on tail entities. We observe that incorporating synthetic data can improve the ASR performance on the entity test set by 0.8% and on the tail entity test set by more than 2%. Future research will explore applying this technique to languages other than German, as was the case for initial experimentation. We also intend to use Text-to-Text (TTS) systems to generate synthetic audios for the synthetically generated textual data for end-to-end ASR model training.

Cloud Provisioning Connector Redesign

Company: Okta
Student: Pengyu Cheng
Industry Supervisor: Stephen Kwo
Academic Supervisor: Eyal de Lara

Okta's mission is to improve the connections between people and tools to make companies more productive and secure. Provisioning is an integral part of the on- and offboarding process and the original Cloud Provisioning Connector (CPC) made it possible for ISVs to build and support provisioning integrations to cloud services that are made available in the OAN (Okta Application Network). However, with the growing amount of integrations being created at Okta, there is a need for a more scalable and lightweight implementation of the CPC Platform that can handle communications (during provisioning) between Okta Service and CPC Platform more efficiently.

The goals of this research project include: (1) find a replacement for the CPC Agent that is not resource-intensive. (2) design a new CPC Platform that can reduce the communication time as well as scaling to all the other applications. By replacing the stand-alone agent with Kafka as the message broker, we are able to reduce the time usage by 60%-75%, which greatly increases the efficiency during user provisioning.

As security is considered a top priority during provisioning, the next steps would involve researching best practices to eliminate the security concern of the newly designed CPC Platform during provisioning.

A framework for supervised and semi-supervised video representation learning

Company: Layer 6 AI
Student: Keyu Long
Industry Supervisor: Maksims Volkov
Academic Supervisor: Animesh Garg

We propose a novel framework to efficiently utilize both labelled and unlabeled data to learn the video representation. Traditional supervised learning methods mainly focus on training a network that achieve the greatest performance on specific task, while self-supervised learning methods aim to learn a generalizable representation for downstream tasks, without any label information of the pre-training set. Our method, however, tries to close the gap between these two approaches. With our specially designed supervised contrastive loss, we managed to learn representations that transfer to downstream tasks well with minimum fine-tuning. We show the state-of-the-art result on multiple popular video datasets, which include Kinetics-400, UCF and HMDB.

Estimating Treatment Effects with Meta-Learners and Their Application to Contextual Bandit

Company: Integrate.ai
Student: Pengyu Cheng
Industry Supervisor: Xichen She
Academic Supervisor: Peter Marbach

Integrate.ai (IAI) is an AI startup that delivers insights that drive meaningful business outcomes using Trusted Signal Network (TSN). The goal of the TSN is to boost one client's first-party model performance by incorporating the signals from the third-party data which lives in other business domains in a privacy-preserving way. Model stacking is one example of a class of ensembling techniques where a new model is trained by combining model predictions from multiple predictive models.
For this project, we focused on one particular predictive model used under the stacking method called Meta-learners. They are one class of machine learning algorithm for estimating the Conditional Average Treatment Effect (CATE). For data sources with treatment variables, the CATEs are potentially more informative signals than the marginal propensity scores. TSN currently only supports S-learner. The objective of this project is to implement other meta-learners such as T-learner and X-learner and integrate them into IAI's platform.

After designing and building T-learner and X-learner, we managed to perform some experiments on the benchmark datasets and observed a significant lift of the contextual bandit's rewards compared to the results obtained using S-learner.

The next steps involve finding ways to improve the number and quality of signal candidates we can extract from datasets with treatment information which can consequently benefit the subsequent knowledge transfer across the Trusted Signal Network.

@UofTMScAC

@UofTMScAC

Master of Science in Applied Computing

U of T – Department of Computer Science

Master of Science in Applied Computing (MScAC) Program
University of Toronto
700 University Avenue - 9th Floor
Toronto, ON M5G 1Z5

mscac@cs.toronto.edu
mscac.utoronto.ca
+1-416-978-5180