Honours and Master of Data Science Mini-Conference

Friday 6th October 2023

Robert Street Lecture Theatre

Programme

10:00 **Welcome: Cara MacNish**

Session 1 Honours **Chair: Roberto Togneri**

10:05-10:30 **Improving the Accuracy and Completeness of Guitar AMT through Advanced Machine Learning**

Declan Ball

Supervisors: Roberto Togneri, Zhanh He

Automatic Music Transcription (AMT) is a growing research area that has been gathering attention due to its potential significance in educational and cultural applications. AMT is the automated conversion of audio signals into formal music notation or sheet music. However, the field's concentration on guitar transcription has been somewhat narrow, and existing research addresses isolated challenges without offering holistic solutions. This gap necessitates the objective of this honours project which is to develop a more accurate and comprehensive guitar AMT system capable of converting an audio clip of guitar music into a tablature file.

Utilizing the GuitarSet dataset, machine learning techniques, and several algorithms, this research aims to combine and improve the existing solutions to each of localised challenges of guitar AMT. The study engages in a comparative analysis of current techniques to identify optimal strategies for both general-purpose and specialized transcription tasks. In particular, we contrast the performance of an enhanced Convolutional Recurrent Neural Network (CRNN) model against conventional cost-function-based methods for note prediction. This investigation thereby contributes to the expanding landscape of AMT by providing a more accurate, efficient, and complete system for transcribing guitar music.

10:30-10:55 **Evaluating Automatic Music Transcription Accuracy with Audio Compression**

Nicholas Chua

Supervisors: Roberto Togneri, Zhanh He

Automatic Music Transcription (AMT) is the process of converting a music signal into readable symbolic notation. It is a rapidly evolving research area of great importance to the music industry, due to its applications in musical education and digital music processing and retrieval. The state-of-the-art MT3 model developed by the Google Brain Team has emerged as a leading AMT solution, being an all-in-one deep learning model able to proficiently identify and transcribe multiple instruments from a singular track. However, there has been minimal exploration in the relationship between transcription accuracy and varied degrees of audio compression. This honours research presents an evaluation of AMT performance under various sample rate compression levels and evaluates the integration of Music Source Separation (MSS) models as a potential front-end for MT3 to tackle its effects. Both monophonic and polyphonic audio signals from the datasets MAESTRO, SUPRA-RW and SLAKH were used. Furthermore, the MSS models Demucs and Spleeter were used alongside MT3 to analyse its capabilities in combating audio compression with and without a dedicated MSS front-end. The insights of this project are important not only for further research in AMT but also for industries reliant on robust audio processing and transmission under varied conditions, especially for cases of extreme audio distortion and compression.

10:55-11:20 **Opinion Dynamics Models in Empirical Networks**

Joel Wildman

Supervisor: Mehwish Nasim

This project concerns the simulation of large scale trends in hypothetical social networks. Opinion Dynamics Models are a family of social models that seek to emulate key features of sentiment dynamics. Modern models feature increasingly complex rules resulting in advanced emergent behaviour but have produced questionably applicable results. Despite great bounds in the scope of model computation and variety, models often fail to ground themselves in observable/established phenomena and thus sacrifice plausibility. A particular area of weakness in this respect is that of social network topology, or the structure of the environment in which the simulation runs. This research intends to assess the effects of changing these environmental conditions, introducing existing models to empirically justified networks to evaluate whether it introduces meaningful variation. The results of this simulation will reveal whether modern, highly complex models are sufficiently susceptible to changes in network topography to justify further investigation into this area.

Results indicate that variations in network topography have notable effects on some models, providing that the simulation rules are not too dominant. Strongly mathematically defined rules see little variation in performance due to their already overwhelming bias, but more complex and less convergence-inclined rules do see noteworthy changes under specific conditions.

11:20-11:35 Break

Session 2 Honours

Chair: Naeha Sharif

11:35-12:00

Multimodal Attention-Based Neural Networks for Prediction of Cognitive Decline

Jamie Vo

Supervisors: Ghulam Mubashar Hassan, Naeha Sharif

Alzheimer's disease (AD) is a neurodegenerative disease characterised by memory impairment and cognitive decline. With the advancement in Artificial Intelligence and Machine Learning algorithms, there has been growing interest its applications in the diagnosis of neurodegenerative diseases such as AD, and its early stage, mild cognitive impairment (MCI). Recently, ensemble based convolutional neural networks (CNN) have been applied to leverage the strengths of different modalities for MCI and AD diagnosis. These approaches include combining multiple modes of imaging data, resulting in improved performances in AD and MCI classification. However, these approaches all are limited by their lack of cross-modal interactions. Different modalities may have varying weights and importance in AD classification, and therefore it is important to determine which modalities and features should be focused on. To this end, we investigate the effectiveness of attention in the fusion of different neuroimages. We build a multimodal attention based CNN which combines both PET and MRI images by using attention based mechanisms to predict the cognitive decline of individuals.

12:00-12:25

Using Object Detection to Improve Medical Report Generation Accuracy

Lachlan Bassi

Supervisors: Naeha Sharif, Ghulam Mubashar Hassan

I will be presenting my research on using object detection to enhance the accuracy of medical report generation. This project had two main components. The first involved training an object detector to identify lesions in Fundus Fluorescence Images. For this, I utilised the labelled lesion data from the FFA-IR dataset, which was curated by Mingjie Li and colleagues in 2021.

In the second phase, I applied the object detector to all the images in the FFA-IR dataset. For images with identified lesions, I cropped them to only showcase the detected lesions. With these refined images, I then tested the accuracy of medical report generation. This new approach's performance was benchmarked against the original baseline, which didn't involve any cropping.

The impetus for this experiment stems from the limited research on the role of object detection in Medical Report Generation. One of the primary reasons fuelling this motivation is the belief that by minimising external noise and centring on the region of interest, the model might yield more accurate medical reports.

12:25-12:50 **Automatic Knowledge Graph Construction and Question Answering using Large Language Models**

Karam Thethy

Supervisors: Wei Liu, Michael Stewart

This work focuses on investigating the use of Large Language Models such as ChatGPT to automatically generate structured data from ordinary written text. This problem is crucial since it is much simpler to organise and interpret structured data than unstructured text. The structured data type investigated is the Knowledge Graph, a structured map of all connected concepts and relationships found in a text. Historically, constructing Knowledge Graphs from text has required a huge amount of training data to teach models how to correctly recognise concepts and relationships. However, with the powerful language understanding abilities of Large Language Models, this requirement is no longer necessary. Using these language models to construct Knowledge Graphs has been done before, however their quality has never been robustly evaluated. This work is the first to use Large Language Models to first generate Knowledge Graphs, and then evaluate their quality by testing them on various question answering datasets. These datasets range from asking simple questions to asking complex, multi-step reasoning questions to test how well the graph captures the essence of the dataset's text. The results are still being gathered as of current, however, any accuracy achieved is noteworthy due to the novelty of the research.

12:50-2:00 *Lunch Break*

Session 3 Honours **Chair: Caren Han**

2:00-2:25 **Transforming Cards Against Humanity: reinforcing style in fill-in-the-blank styled humour predictions**

Daivik Anil

Supervisor: Tim French

Throughout time, jokes and humour have been used to strengthen the bonds of people and the means in which they communicate. Large language models (LLMs), such as GPT-4, have been utilised in the recognition and generation of humour to varying results. However, context-specific humour remains under-explored in computational research due to its ambiguity and complexity. Cards Against Humanity, a fill-in-the-blank style game, provides for a perfect canvas to explore this, due to the game's inherent focus on shocking and dark humour. Analysing the medium can provide useful insights into the mechanisms that underly humour, and specifically AI-driven humour.

This research focuses on applying BERT language models to the task of rating and predicting winning jokes in simulated rounds of Cards Against Humanity. A large dataset of 300,000 rounds is used to explore this task. To tackle the challenge of the subjectivity of humour this research implements and applies proximal policy optimisation (PPO) to optimise for a darker, more offensive style of humour. In this seminar, we will compare and analyse the influence of PPO on a fine-tuned model's rankings of offensive humour. We will also qualitatively analyse the importance the models puts on different features of the jokes.

2:25-2:50 **Scaling Active Inference**

Fraser Paterson

Supervisor: Tim French

Active Inference is an emerging, first principles account of brain function, adaptivity and Intelligence. Practically speaking, Active Inference unifies the operation of action and perception under a single imperative: free energy minimisation. In this respect the theory is a highly parsimonious account of agent-based ecological adaptivity. The theory has made impressive strides in recent decades in accounting for a wide range of classically "intelligent" behaviour. Indeed, it is increasingly plausible that Active Inference might constitute a fundamental mechanism of Intelligence per se. Thus far, artificial instantiations of these agents have been relatively limited to hand-crafted models, operating on small, discrete state/action spaces, and usually for the purpose of simulating a specific cognitive phenomena. Implementations capable of operating in larger, more complicated environments are currently hamstrung by issues of scalability, and computational tractability. This thesis has sought to address these dual issues by means of Deep Learning and Heuristic Tree Search, in an attempt to afford the operation of Active Inference for increasingly "real world" problem instances. The resulting scheme is applied to various environments in the "Gymnasium" package - formerly OpenAI gym - and a host of comparisons are made between our implementation and the more established "Reinforcement Learning" paradigm.

2:50-3:15 **Multi-modal Multi-task Game Situation Understanding**

Shan Ng

Supervisor: Caren Han

Esports, a competitive event where skilled players and teams compete professionally in gaming tournaments for a prize pool. These events are live streamed on online platforms, allowing people to watch from different corners of the world. However, due to the inherent complexity nature of the game, it can be difficult for newcomers to comprehend what the event entails. The chaotic nature of online chat, the fast-paced speech of game commentary, as well as game-specific user interface further compound the difficulty for users in comprehending the gameplay. As such, it becomes imperative to provide concise summaries of the game's situation, facilitating a more accessible understanding of the competition's progression for newcomers. This task can be seen as multi-modal multi-task learning, where multiple streams of information from the livestream platform are processed separately before they are integrated together to form a joint understanding model. In this research, we propose a model that integrates real-time chat interactions alongside livestream information, with the ultimate goal of constructing a framework that enhances the comprehension of the ongoing game situation. By combining these disparate streams of information, our study aims to contribute to the advancement of viewer experience and engagement in the field of livestream games.

3:15-3:30 *Break*

Session 4
MDS

Chair: Débora Corrêa

3:30-3:50

Optimising Information Gathering Following a Major Crime Incident

Mitchell Doody-Burras

Supervisors: Shannon Algar, Michael Small

A role of the police is to respond to crime incidents, and in most cases, they prioritise the speed of the response. However, when responding to major crime incidents, gathering information may be equally important. While first responders reach the incident, following responders should balance information gathering with timely arrivals. Major crime incidents are larger and last longer, making information gathering an important component to resolve these incidents. To increase information gathering, we optimise the routing of police responders. Our approach employs computational modelling to simulate complex responses to major crime incidents. A combination of an inverted ant colony and shortest path algorithm is used, whereby routes taken deter following responders from taking that same route, effectively forcing information gathering along alternate routes. It allows for the exploration of interactions between individual responders and their environment, which collectively form the response. These simulations encompass various scenarios and response strategies, enabling the estimation and better understanding of response outcomes. The results show solutions that make the most use of limited police resources to maximise incident coverage whilst still ensuring a fast response time. These models equip law enforcement with a means to expedite resolutions for such challenging incidents.

3:50 *General Session Close*

3:55-4:15

Closed Session (CEED)

Symbolic Time Series Analysis for Characterising Dynamics in Industrial Machinery

Omri Ram

Supervisors: Débora Corrêa, Rachel Cardell-Oliver, Charlie Musca (RCT), Kevin Winchester (RCT)

Decreasing costs and increasing availability of sensors and computing power allowed wide adoption of the Industrial Internet of Things enabled predictive maintenance. Challenges are still posed by data collected in an industrial setting which is often unlabeled. Despite the accrued attention, there are still not many works detailing unsupervised approaches to recognize degraded performance. This study seeks to explore the use of a D-Markov chain constructed on symbolized time series data to identify time series segments exhibiting degraded performance in industrial machinery. To date, the research has established that the methodology is capable of identifying time series segments corresponding to degraded performance on a benchmark dataset. Further assessment by domain experts is still necessary to verify that the method works on real world data.

4:15 *Conference Close*