Technical Report

Course Number: CS 4982

Summary of Science Atlantic Mathematics, Statistics and Computer Science Conference

Name: Eric Foote

ID Number: 3502094

Monday October 25 2021 Supervisor: Dr. P.Mahanti

Department of Computer Science

University of New Brunswick Saint John

1. The name of the conference is Science Atlantic Mathematics, Statistics and Computer Science Conference 2021
2. The dates of the conference was Friday October 22 to Sunday October 24
3. I was able to attend 2/4 of the plenary talks
4. The first plenary talk I was able to attend was the Fields Lecture by Dr. Luke Bornn, which was titled “From Pixels to Points” and the second was the Industry Lecture by Dr. Nithum Thain, which was titled “Machine Learning Fairness”

Each talk summary is on its own seperate page

Summary of “From Pixels to Points”

The first talk I attended was the first on Friday October 22 and it was titled “From Pixels to Points”; it was the Fields Lecture by Dr. Luke Bornn. The lecture covered two areas of work one for basketball and the other for football(soccer).

The first is to understand NBA (National Basketball Association) offences. The work that was done was to help basketball teams solve the internal problem of tagging video to help the coaching staff answer questions about specific teams offensive strategies. This is a clustering task similar to document analysis. For example, “find all the spurs hammer sets from the entire season” is similar to “find all instances of the phrase white house in a given set of documents”. The goal is to be able to model the possession structure of each teams offences. The first method that was introduced was have a coach hand annotate each individual possession and then feed that into text analysis models; this is an extremely hard thing to do since there are 1000’s of these individual offensive plays through a season. The team then used player-tracking data to turn plays into a document which can then be fed into a text analysis model. The methodology the team used was for each individual player we take their movements and slice it up into individual segments; then we cluster each individual player into 250 trajectory cluster (each cluster is a function through time) then to turn it into a document we use a topic model: LDA (Latent Dirichlet Allocation (Blei 2003)) to model possession. Another use of this data was to examine all instances of a similar play. In summary, this project was to organize a database of hundreds of thousands of NBA possessions by offensive structure, through the use of LDA and functional clustering which allows coaches to do deeper dives into different plays.

The second field of work covered was to look at expected value for possession in football using data from the Spanish League. This problem was solved earlier through the work done on NBA possession data to look at expected value of possession. One key challenge though was though is football possession is nuanced, passes can go anywhere, here context is critical. The goal of the project was to solve the expected goal or goal against given all spatiotemporal information until time t. In other words, we need to model the EPV (expected possession value) as a Markov Decision Process. The team needed to add context data to the possession data. This ended up being a big complicated neural network that was broken down to three components due to the laws of expected value: pass, ball drive and shot. This allowed teams to do decision-making analysis through layers of probabilistic surfaces.

Summary of "Machine Learning Fairness"

The second talk I was able to attend on Saturday October 23 was the Industry Lecture by Dr. Nithum Thain, which was titled "Machine Learning Fairness". The goal of the talk is what is machine learning? How to measure its fairness? Why do fairness challenges arise and what can be do about them? It began with a look at the COMPAS system for criminal offences used in the USA Justice System; it was a series of questions that labelled people as either high or low risk. It was determined in 2016 to be racially biased. Then it was discussed that there is no single definition of artificial intelligence that is universally accepted by practitioners; while machine learning has a universally accepted definition of using statistical techniques to give computer systems the ability to "learn" with data, without being explicitly programmed. It is also used to extract patterns from the data. A typical machine learning pipeline was outlined of collecting training data, then training the model; following that evaluating the model and finally, deciding whether to deploy. One key thing is that when you collect training data ask the question where potential biases can enter the data. When we tweak the model we use calculus if we feel the labelled data when added into the model gave us a bad answer. To evaluate the model we consider two factors: recall (i.e. if you are sick, how likely are you to be labelled as sick) and precision (i.e. if you are labelled as sick, how likely are you to be sick). Once we have trained and tweaked the model we need to look at its fairness. One thing to consider is, any number of human bias can creep in at any step, how it was collected, how it was labelled, what model was chosen, how it was trained, etc… A question arises though of what does it mean for a model to be fair? Fair to whom? One approach is to compare the precision/recall across predefined groups. One thing to remember though is it is mathematically impossible to balance both precision and recall, as you improve one it makes the other worse; you have to choose one or the other. When revisiting COMPAS, we see that when you look at the fairness through the lenses of recall, similar to the press, it is indeed biased; however, when you look at it through the lenses of precision, the algorithm is fair. Both cases are correct. One thing to get from this talk therefore, is that there are more questions than answers. How will my decisions affect users? How can I improve my model training pipeline? What are the gaps in my datasets? Does my model pick up spurious correlations? Should I ask for user demographics? Is my evaluation comprehensive? When should my model not be used? What isn't captured by statistical notions of fairness? Furthermore, to improve fairness consider three possible steps: dataset augmentation: add better examples from underrepresented groups. Adversarial modelling: train model to be unable to predict demographic attributes. Finally, user interface adjustments.