

CSCI 600
Big Ideas in Computer Science

Final Report
Spring 2021

OMER BIN ALI BAJUBAIR
Z1905006

INTRODUCTION

CSCI 600 (Big ideas in Computer Science) is a 1 credit course offered by the Computer Science department so that we can get insights of what research our Professors are doing and what projects they are working on. We were given Guest lectures by our professors of the CS department in their area of interests. I found the lectures very Insightful. There were wide variety of research topics presented in front of us ranging from Bioinformatics, Artificial Intelligence, Data Science, Medical Imaging, Data visualization, High Performance Computing and Machine Learning.

I was amazed with the depth of research that is going on in our department and have even asked questions at the end of the presentations to clear my doubts. This motivated me to take up research under the professors so that I can contribute and learn from them. After all this course gave me insights on how vast the computer science field is, and it is not just limited to programming. For each lecture I have written my views and takeaways in this report. Attached is the Schedule of the lectures we had.

- Mar. 25 - Dr. Jie Zhou - NIU
- Apr. 1 - Dr. Reva Freedman - NIU
- Apr. 6 - Dr. Hamed Alhoori - NIU
- Apr. 8 - Dr. Kirk Duffin - NIU
- Apr. 13 - Dr. David Koop - NIU
- Apr. 15 - Dr. Mike Papka - NIU / Argonne National Laboratory
- Apr. 22 - Dr. Pratoool Bharti - NIU

Lecture 1:
Speaker: Dr. Jie Zhou
Date: 25th of March 2021
Topic: Learning-based Bioimage Informatics

Dr. Jie Zhou is a Professor at NIU. She gave a presentation on the topic “Learning Based Bioimage Informatics” which was based on Machine Learning and Bioimage Informatics. Firstly, she explained the steps involved which are counting, reconstruction, annotation and behavior analysis. And also, she presented some of her research which includes the visualization of dendritic cells, cell counting in higher dimensions. She further explained about a topic called Discriminative Nuclei Detection and gave the algorithmic flow for it which was a method to overcome the challenge where the cells have dense clusters. I got curious and asked her the question: What tools/ Packages she uses for this analysis? To which she replied she uses ImageJ which is a platform to analyze the cell processing and BIOCAT (Bioimage Classification and Annotation Tool). For reconstruction she uses C++ / Java and for annotation MATLAB and for Behavior Analysis she uses python. To estimate the dendritic length based on tracing she uses Graph Traversal Algorithms like Dijkstra Algorithm for finding shortest path between two nodes and divide and conquer techniques.

The research area she works on is very fascinating because it breaks all the norms where people think computer science is a field in which programming is only done to build apps, but I think in this current pandemic situation of covid 19 it can be used to understand different mutations which can be beneficial to the mankind.

Lecture 2:
Speaker: Dr. Reva Freedman

Date: 1st of April 2021

Topic: Applications of Artificial Intelligence

Dr. Reva Freedman gave an introduction of what is Artificial Intelligence and their applications. Her research area is the Centre of Intersection of AI and education research and NLP. She spoke about the simulation of brain. She gave various applications of AI which included Google Maps that uses A* algorithm (one of the most popular Graph traversal Algorithm) On a day-to-day basis we use the google maps, but we are unaware how it works behind the hood. She also gave the example of Google Translator which uses Deep Neural Networks and most importantly Recommender System, which is used in Social Media, ecommerce etc. Common example we can see is if we buy something from amazon it will notify us that the persons also bought these items. This is based on AI. Her research area is very diverse in the field of Intelligent Interactive Systems (IIS) and Artificial Intelligence in Education (AIED) and CRAM (cardiovascular anatomy map). I was amazed with the vast applications that Artificial Intelligence possesses in our daily life. This lecture helped me better understand things like the recommendation system and I got a thought that we can build a system with AI which can recommend people specific medicines based on the disease they have without human intervention.

Lecture 3:
Speaker: Dr. Hamed Alhoori

Date: 6th of April 2021

Topic: Rethinking Science through Data Science

Dr. Alhoori gave presentation on the topic “Rethinking Science through Data Science” which included the use of ML to study the broader impact of research. First, he explained understanding the complex Information Behavior then he went on to explain the characteristics of good science which consisted Systematic, Rigorous, Replicable, Novel, Ethical and Generalizable in a concise manner. Then he went on to explain Altmetric with a comparison of Citation and Altmetric in social media. He discussed how Altmetric can be used to predict citations for scholarly references and predict the citation counts per article. I found this lecture very fascinating as I am an avid reader of Google’s scholar articles and how we can use Machine learning to do patent citations. He also provided prospective on how to identify specialized research community and researchers changing interests. Having done some projects in Data Science I can correlate with the things he was saying and how impactful data can be in our lives. I believe there can be better ways of telling the story of research outputs than just pointing to them and the journal they are published in using Altmetrics.

Lecture 4:
Speaker: Dr. Kirk Duffin

Date: 8th of April 2021

Topic: Medical Imaging, Fluid Simulation, Representing Geographic Information in Compact Way

Dr. Duffin gave presentation on multiple topics ranging from Fluid simulation, Medical Imaging and data map representation in a compact way. The topics were very interesting as I had no previous idea on how the Graphics are used in MRI and X-ray scans. He also explained about the problem of protons which is a charged particle that do not travel in a straight line and Proton Imaging matrix which is very big and requires big computing power. When asked about the time complexity to which he replied it depends on how sparse the matrix is. He has been in this research for more than a decade and doing some amazing work. He explained about Proton Cancer Therapy and how using Computer Graphics we can come up with an image to visualize the impact of a proton beam on a human body with Ionizing radiation and Bragg peak for radiation imaging. Then he went on to explain the Fluid simulation and symmetry in design. One of the things which caught my attention was the representation of geographic information in a compact way as the data structure used in that was quad tree which is like binary tree. Having Interest in data structures and algorithm I could visualize it with comparison to Different graph algorithms like Dijkstra and A*. He explained this by giving example of map of Illinois and city of Chicago and its adjacent states and how each one of us can interpret and represent the edges and boundaries and that how important are the edges and vertices in the field of Computer Graphics. He also explained

Wave craft which showed 3D weaving. I found this idea also very insightful as this can be used to create fancy clothing stuff very easily. I got ideas where we can use Computer Graphics to make more engaging educational video games so that children can play and learn parallelly.

Lecture 5:

Speaker: Dr. David Koop

Date: 13th of April 2021

Topic: Visually Summarizing Dense Datasets Level by Level

Dr. Koop gave a presentation on summarizing dense data and data visualizations. He explained about Simprov which is for tracking provenance in web applications which denotes how we got to a particular result. The interesting topic that the speaker presented was that of ship travel path where he could visualize the data and present the path of ship travel in more clear way. He explained how Notebook Output representations can be improved by removing ambiguities and anomalies in data. He also focused on the steps that are required for the data visualization. One of the questions that I asked was how we can make data cleaner so that we can get better visualizations to which Dr. Koop replied there are various ways to do this we can either remove the outliers by using techniques like skipping the missing data or apply filters. I feel that data visualization is a key component in understanding data. There are a lot of problems while dealing with data which can be huge volume of data, overlapping, or missing data. Having tons of data with us it is completely useless if we cannot understand the data and study the hidden patterns in it that is what he explained in his presentation. Visualization is about getting Insights from data which Dr.

Koop proved through his presentation. He showed the Interactive approach, which was overview first, then zoom-in then filter and at last details on demand. I found his overall presentation were interesting and informative as I have already worked on Jupyter notebooks and I was able to Correlate. He also explained about workflow differences and google page rank algorithm. This lecture was very Insightful for me one idea that I can think of for applying the visualization is to show the path that the marine species take in the sea and oceans so that we can know the migration pattern and where they can be kept safer away from harmful environment and poachers.

Lecture 6:

Speaker: Dr. Mike Papka

Date: 15th of April 2021

Topic: Big Ideas from The Data, Devices and Interaction Lab

Dr. Papka primary research interest is in the field of Information Visualization, Scientific Visualization and Analysis, Augmented Reality and Virtual Reality and High-Performance Computation. He has worked at Fermi National Accelerator Laboratory and Argonne National Laboratory and said how he shifted from physics to Computer science due to his passion. But his talk was mainly focused on High performance computing, Internet of Things and Augmented reality / Virtual reality. Having myself worked with supercomputers for nearly 2 years at NTT DATA I could corelate the things he was saying.

Supercomputers like Cray X-MP are heavy duty computers capable of operating at a high-performance speed.

He seems excited about the new supercomputer that Argonne National Laboratory is going to get by 2022/2023 named Aurora. When I asked 2 questions at the end of the class Dr. Papka gave a very detailed answer. My questions were in mainframes the jobs are submitted to a job queue then the resource scheduler decides which job to execute? is it the same process for HPC to which Dr. Papka replied yes, it is the same process for HPC, and it is a vast research topic to learn about the resource scheduling. Another question was what if a node fails will it impact the other running nodes. He replied that until it is a very bad failure it will not impact the other cluster nodes, and this is also a very interesting topic which is fault tolerance which is used to recover from failure. I found Dr. Papka possess deep knowledge in every aspect of the research. Even he explained about Array of things which is based on Internet of Things. Sensors are placed in the city of Chicago to better understand the traffic movement, Pedestrian movement so that they can make the life of people better by analyzing the data. With a boom in the data technology there will be more need of HPC in the future I feel, and we will get a chance to work on it in every aspect.

Lecture 7:

Speaker: Dr. Pratoool Bharti

Date: 22nd of April 2021

Topic: MLOps : From Model-Centric to Data-Centric AI

This lecture was supposed to be given by Dr. Pratoool but due to his unavailability we were shown a recorded video by Andrew NG who is a computer scientist and founder of Google Brain. In this lecture Andrew

explains MLOps from model centric to data centric AI. Andrew explains the AI system as the combination of code and data. He also gave an example of inspecting steel sheets for defects where we can identify the defects using computer vision. He also elaborated on the problem whether the team should improve the data or the code. By using data centric approach, we were getting more accuracy over model centric approach. He also gave an analogy where Data is food for AI. He also explains that he has reviewed 100 abstract research papers out of which 99 focused on the ML models and one talked about the data which he compares with the 80/20 rule. Training the model is only a small part of a big machine learning project. He also explains the lifecycle of the ML project which consisted of the steps scope project, collect data, train model and deploy in production. We were also shown a Iguana detection example which identifies the species in an image. According to him even if a dataset is small and clean then it considered good dataset compared to big dataset which is noisy. He went on to explain speech recognition to which I was very amazed how ML can be used to make the results more accurate. I got to know the difference between Traditional software and AI software where the latter is interactive. The presentation was ended with a QnA session. What I felt was one on one interaction was missing and even the caliber of the research was much higher. I found his video very informative, and it was a privilege to hear his views from such a great personality who leads the AI team at Google. I was able to understand that how quickly ML has become a very important branch of Computer Science and how for anything and everything ML is being used.