JU CHEN

(201) · 680 · 1987 ⋄ jc7439@nyu.edu

55 River Drive South Apartment 1002 · Jersey City · New Jersey

EDUCATION

New York University *May 2019*

M.S. in Mathematics

Coursework include Stochastic Calculus, Partial Differential Equation, Measure Theory, Numerical Methods, Advanced Topics in Monte Carlo, Machine Learning,Deep Learning, Big Data and other graduate level math and data science courses

Overall GPA:3.64

Purdue University *May 2015*

B.S. in Mathematics Honors & Applied Mathematics, Minor in Economics

Got Senior Achievement Award; Served in the Boiler Out Volunteer Program for four years; Took three graduate level courses

Overall GPA: 3.70

RESEARCH EXPERIENCE

Detecting outlier in a large data set (Algorithm design, Data Mining)

· Defined distance based outlier inspired by Distance-based Outliers: Algorithms and Applications

· Implemented both index-based algorithm and space-based algorithm for low dimension data (to find the outliers defined above) Applied ArrayList in PySpark to store nearby data points in the same cell

· Observed the problem of dimension curse while building that cell and improved the result by Optimize the definition of outliers (by comparing the K-th nearest neighbor for every points)

· Introduced pruning and Birch clustering for efficiency and reduced 50 % of running time. Specifically,used min-heap and max-heap for pruning, R*-tree for Birch clustering (more efficient than k-means clustering).

· Applied this algorithm to the NBA player statistics to find either All-star play or players that have specific strength in one or two statistics. Checked accuracy by visualize the data using R(ggplot2)

Inplementation of Multi-Level Monte Carlo Method Simulation on SDE (Numerically Method, Simulation)

· Implemented the Multi-Level Monte Carlo Algorithm. Aim to minimize the cost of simulation with given accuracy.

· Chose a number L and discretized SDE by applying Euler-Maruyama discretisation into L steps.Break the SDE into L telescoping sums of expectations so it is easy to simulate

· Simulated discretized SDE for telescoping sums and found the optimal numbers of simulation for each of them

· Checked the convergence condition to see if more steps need to be divided.

· Used Matlab to implemented such Algorithm to simulated the trajectories of Black-Scholes formula and found the payoff of European and American options.Estimated the upper and lower bound of their payoffs graphically and numerically.

· Presented the project for our team in front of the professor and students

WORK EXPERIENCE

Data Scientist Intern at Rockpapr April 2018 - Present

· Applied Beautiful Soup for web-scraping to collect raw data of various products.Cleaned and formatted data through Python and SQL so that it can be more easily analyzed

· Inplemented NLP package (NLTK) for name and identity matching in order to match payment information (from bank statement) to corresponding products.

· Optimized the matching result by applying FuzzBuzz package and Logistic regression to get over 90% of accuracy

· Self designed model based on the payment information to detect recursive payments and to predict the next payment

· Compared payment ids associated different periods to make sure only the smallest period is detected

· Collaborated with software engineers to upload the model to flask so it can be tested and used in production environment

PROJECT EXPERIENCE

Mobile Games A/B Testing with Cookie Cats

· Aimed to determine if the first gate in Cookie Cats should be moved from level 30 to level 40.

· Chose 1-day retention rate and 7 -day retention as Key Performance Indicators (KPI).

· Implemented statistical formula to determine the minimum number of samples to be collected. Ran the experiment for two weeks

· Applied hypothesis testing to determine whether the difference is significant.Implemented bootstrap when sample size is limited

Movie recommender system (Item collaborative filtering)

· Implemented Hadoop MapRduce to group data by user, build co-variance matrix and to coduct matrix multiplication

· Predicted user's rating towards movies they have not seen and make recommandations based on prediction

· Conducted one additional MapReduce job to replace the non-rated movie score by the average of user rating towards other movies

TECHNICAL STRENGTHS

Programming Languages	Python, Java, R, SQL
Mathematical Skills	PDE, ODE, Numerical Analysis, Linear Algebra, Probability and Statistics
Other Skills	Data Science, Machine Learning, Microsoft Excel,Linux,Git