DISSERTATION PROPOSAL

Nikhil Malik

"Applications and Economic Impact of Machine Learning and Blockchains"

Monday, November 11, 2019
1:00 pm
Tepper Quad 4219

Machine Learning (ML) Algorithms and Blockchains have been two major technology disruptions in the last decade. On the one extreme, Blockchain attempts to decentralize decision making power to a crowd of anonymous participants. On the other extreme, ML algorithms often centralize decision making into uninterpretable algorithms. The first chapter of my dissertation focusses on an ML application to solve a pressing question on source of attractiveness bias in labor markets. The second and third chapter focus on interaction of market participant with Blockchain and ML algorithms respectively, as well as suggestions for more efficient outcomes.

In the first chapter, we apply Machine Learning techniques to study bias arising from attractive face appearance. We investigate two sources of bias, namely, preference-based and belief-based. Belief based bias against subjects exists because evaluators have group-level priors based on the subjects' attractiveness. These priors are overcome as the evaluator obtains objective signals of performance. Preference based bias exists because evaluators have an inherent taste for a social, romantic or marital relationship with attractive subjects. We use one of the largest archival longitudinal data sets (19,893 MBA graduates) in this area of research to identify these two sources. We find that attractiveness bias leads to a 2.6%-3.6% gap over a 15-year career period. This gap is a result of a preference bias that creates an attractiveness gap of 0.64% per year. On the other hand, belief bias has no significant role in post-MBA professional careers. This is a significant finding because belief bias toward an individual can be minimized by the individuals' performance information. However, preference based biases are much harder to remove. Our empirical setting (20,000 MBA graduate profiles) presents two key challenges in working with unstructured data. First, for an individual, we observe only one current picture, which is taken up to 25 years after the start of the individual's professional careers. We combine a generative deep learning model meant to morph individual's face over age with an interpretable linear model of attractiveness evolution. This allows us to project unobserved heterogeneous evolution of an individuals attractiveness. Second, individuals move across job profiles, companies and locations, thereby making it difficult to directly compare their career growth. We construct a preference order for over 100,000 unique jobs using page ranking of vectorized job text representation.

In the second chapter, we study economic limits on scalability of Bitcoin Blockchain. Its ledger entries grow by addition of blocks of ~2000 transaction every 10 minutes. One would expect that increasing the block size would scale Bitcoin. But, we show that a Block size increase encourage miners to tacitly collude – artificially reversing back the capacity via strategically adding partially filled blocks in order to extract excess economic rents. This strategic partial filling is sustained if the computing power of the smallest colluding miner and cumulative power of the colluding group is larger than a threshold. We show that an intervention banning large miners can breach this threshold thus eliminating collusion. But, this makes the system less secure as double spend attacks become attractive in relation to mining revenues. Owing to this dual threat of collusion and double spend attacks, its untenable to offer a high capacity ledger to users with

widely different willingness to pay fees, bear delay and risk attacks. In fact we advocate miner collusion as a useful mechanism where a chunk of excess revenue raised via sacrificing capacity, endogenously spillover into an investment into platform's security.

In the third chapter, we study the economic impact of Machine Learning predictions in markets with uncertainty. Consider a setting where risk averse market participants routinely make decisions under price uncertainty thus leading to surplus loss. Theoretically, a Machine Learning model with ability to crunch historical data can reduce this price uncertainty by learning an approximate map from product features to price. One would expect the model to make biased and noisy predictions initially but refine over time using new data observations. Unfortunately, the new data observations are – (a) systematically oversampled from feature space where the algorithmic predictions had low noise to begin with and (b) correlated with the algorithmic biased mis-predictions in the past. Effectively the algorithm moves the market to a pattern of new trading activity which makes its own performance look good. The situation is exacerbated because human decisions are influenced by noisy machine predictions and machine's online learning relies on noisy human decisions in a closed loop. We want to empirically validate this phenomenon for Zillow's house price prediction algorithm (Zestimate) and its impact on subsequent buyer-seller house sale decisions. Further, we intend to structurally estimate Machine's learning and prediction procedure as well as human's house buy-sell decision models. This would allow us to study counterfactuals parameter settings where the correlated machine-human feedback loop leads to housing price bubble. Finally, we want to compare a new "feedback aware ML predictive model" against an unaware model in terms of firm revenues, consumer surplus, and robustness to housing bubbles.