NBER Entrepreneurship Workshop 2022

Contact information, including email address.

- Name: Hong-Yi TuYe
- Email: hytuye@mit.edu
- Phone: 8053773328
- Recommender(s): John Horton and Scott Stern

Background

A short description of the course of graduate study pursued, including major and minor fields of study, and the years of enrollment at relevant institutions.

- Third year PhD at MIT Sloan.
- Primary field: Economics of Technology and Digitization (Information Technologies Group)
- Secondary field: Technological Innovation, Entrepreneurship, and Strategy (TIES)

Courses (Current ones are flagged)

- MIT Economics: Econometrics and Microeconomics PhD Sequence, Industrial Organization Sequence,
  - Econometrics PhD Sequence
  - Microeconomics PhD Sequence
  - Industrial Organization Sequence
  - **High dimensional causal inference with ML (Victor Chernozhukov, Current)**
- MIT Sloan
  - Seminar in the economics of technology and digitization (Erik Brynjolfsson)
  - Seminar in the sociology of technology (Wanda Orlikowski)
  - Seminar in the economics of innovation and entrepreneurship (Scott Stern and Pierre Azoulay)
  - **Seminar in experimental design and causal inference (Dean Eckles, Current)**
  - **Seminar in corporate and entrepreneurial finance (Current)**
- Other MIT:
  - Advanced Natural Language Processing (EECS)
  - Mathematical Statistics (Math)

Research Interests

Overview

I am primarily interested in how competition and innovation policy needs to be revamped for the digital economy. The first topic is directly relevant to this workshop whereas the second and third are digitization topics I've worked on in the past or currently working on.

Measuring Strategy with Learned Embeddings

- Goal is to gather 10-K documents from the SEC, convert to textual data and create a firm-level embedding so that we can cluster the documents. First the clustering will be guided by established S&P sectors. Eventually the goal is for the algorithm to learn new groups that might not seem obvious to the S&P's classification but could signal changes in product strategy under the surface. The reason we can only make claims about product strategy is because we are focusing on the "business overview" and "risk factors" chapter of the 10-K. This often ends up looking at the competitive landscape through the lens of product markets and customer segmentation.

- Ultimately, I would like to learn an embedding space for public firm strategies via websites and financial statements in order to quantify strategic and product differentiation over time. In management and economics research, scholars have lagged in coming up with a measurement of strategy that leverages firm communication to markets. How the firm markets itself and how it sees its current risks can signal shifts in latent strategies that might be captured by state of the art NLP approaches.

- The most recent attempt in this research agenda comes from the management and strategy literature. Guzman and Li (2019) obtained data on all companies available in Crunchbase founded between 2003 and 2019 that have a website. Hoberg and Phillips (2016) is another key paper in this line of inquiry.

- In my project, I show that contextual embeddings (BERT-like models) can almost double the performance of our clustering algorithms when compared to LSA-TFIDF and word2vec. As such, it would be interesting to see how these embeddings affect the firm differentiation scores proposed by Hoberg and Phillips (2016) and Guzman and Li (2019).

- Assuming that we have a validated set of embeddings for describing companies, my plan is to address some of the following questions of interest to economists working entrepreneurship and strategy:

  - How do industry distances evolve before and after two firms from two different industries merge? For example, when Amazon purchased Whole Foods, did that spur interest in companies that share a similar space to Amazon (e.g. e-commerce) to consider similar M&As or expanding product and service offerings as part of a catch-up or (fear of missing out) strategy?
  - Can we relate changes in job postings to changes in strategy? This might be a little bit more involved in terms of data because it would require another embedding for job postings and detecting sudden shifts across both sets of embeddings. E.g. one could imagine Apple increasing its hiring of chip designers and engineers in the years leading up to its M1 chips. Any sort of significant signal in this exercise can also serve as validation of learned embeddings.
  - "Kill zones" and ex-ante innovation: We can apply these embedding techniques to other related datasets like business registration information or application descriptions on the app stores (Apple/Google). Then we can derive a richer representation of how much ex-ante innovation exists (especially in the target firm's industry) before a large M&A deal occurs. Having a firm differentiation score will help us pin

down the set of all firms playing in the space before M&A and after the M&A and the investment flows associated with these firms. The study would first seek to corroborate or refute the results found in Kamepalli et al. (2020). An additional extension is to include pitch deck data from accelerators like Y-Combinator and scrape through these pitch decks to measure entrepreneurial interest in spaces affected or adjacent to the M&A space. Pitch decks can be transformed into text and processed with the same NLP pipeline that we use for 10-Ks and job listings. Constructing a dataset of pitch decks that is representative of global or American VC dealflow (potential and realized) would allow us to come up with a measure of entrepreneurial interest.

- This final question is the most relevant one for the entrepreneurship workshop. I will use the workshop as an opportunity to catch up to the research frontier and see what other gaps can be filled with this firm differentiation measurement. Since entrepreneurship papers often focus on smaller private firms, we can rely on Guzman and Li (2019)'s dataset from the Wayback Machine as well as hoping to get access to a dataset of startup pitches from an accelerator in order to capture the space of non-public companies and startups.

Algorithmic Collusion

- Solo-authored project, paused, pending finding a convincing empirical setting to test out algorithmic collusion theories

- Traditionally, antitrust authorities have placed very little emphasis on policing tacit collusion (i.e. collusion amongst firms without explicit communication/coordination). Firstly, lack of coordination makes successful tacit collusion very difficult, especially when there are several players and economic shocks. Secondly, even if firms were to successfully engage in tacit collusion it would be hard to detect. There is a growing literature experimenting with whether and how algorithms can collude under different economic scenarios that can start to inform the debate over algorithmic collusion. In my paper, I seek to extend the work done by Calvano et al. (2020) and Klein (2018) in two key aspects — one methodological and one conceptual. Both papers acknowledge the limitations of using tabular Q-learning since it pre-dates deep reinforcement learning techniques and today's state-of-the-art (SOTA) algorithms in model-free reinforcement learning (RL) typically make use of deep learning to approximate the value and policy functions. Therefore, I apply some basic deep RL approaches (like DQN and Double DQN) to the same set of questions that Calvano et al. and Klein ask. Conceptually, I am also interested in looking at the long-run behavior strategies when different types of learning agents are involved since in reality it is unlikely that all firms will deploy the same type of RL agents.

- I was able to corroborate Calvano et al.'s finding that price collusion is possible between two tabular Q agents, but I also showed that the dimensionality of the action space is a key caveat to the result. I found that as action spaces increased in size, the tabular Q agents struggled to find the collusive strategies. The more sophisticated DQN agents also struggled to collude, and surprisingly (at least compared to my own priors regarding algorithmic complexity and likelihoood to collude) their eventual move towards Nash-Bertrand prices occurred in both complex and simple action spaces. Finally, I showed in asymmetric games how the more sophisticated algorithms could achieve better outcomes than the less sophisticated agents

Covid-19 and Remote Work

- The first paper is co-authored with Erik Brynjolfsson, John Horton, Adam Ozimek, Daniel Rock, and Garima Sharma.

- We use a large-scale, nationally-representative online survey (Google Consumer Surveys) to track its evolution across demographic groups, geography, and over time in the United States. We ran the survey in four separate waves in April 2020, May 2020, July 2020, and December 2020, covering a total of 80k+ respondents. By May, over half of Americans were working from home, more than tripling the pre-Covid fraction of remote workers. Younger workers (below 35 years) and women were more likely to switch to working from home. There was also substantial geographic variation, with larger changes in the Northeast (38 percent) than the rest of the country (30 percent). Across states, places with a high pre-Covid share of workers in occupations suited for remote work experience the largest switches into it. Starting in July and continuing into December, there was a slow shift back to commuting. This shift was also smaller for places that had a higher share of workers in remotable work. Taken together these results provide evidence of hysteresis as the pandemic accelerated learning about which jobs can effectively be made remote.

- In our second paper, we look to expanding the time and question scope of our surveys through a collaboration with Gallup. The goal is to deploy two survey waves in order to capture the longer term effects of Covid after the world has returned to a more normal state. The first wave as already been deployed and we are fielding the second wave in January 2022. The goal of the second wave is to measure remote work persistence after social distancing and mask mandates have been lifted both nationally and across all states.

References

Guzman, Jorge and Aishen Li (2019), "Measuring founding strategy." URL https://doi.org/10.2139/ssrn.3489585.

Hoberg, Gerard and Gordon Phillips (2016), "Text-based network industries and endogenous product differentiation." *Journal of Political Economy*, 124, 1423–1465.

Kamepalli, Sai Krishna, Raghuram G. Rajan, and Luigi Zingales (2020), "Kill Zone." Working Papers 2020-19, Becker Friedman Institute for Research In Economics, URL https://ideas.repec.org/p/bfi/wpaper/2020-19.html.

Hong-Yi Tu Ye

Updated December 2021
hytuye@mit.edu

EDUCATION

Ph.D. Candidate at Sloan School of Management — Cambridge, MA
Massachusetts Institute of Technology — Aug. 2019 – Present

- Area Focus: Economics of Technology, Innovation Strategy
- Ph.D. Minor: Statistics
- Graduate Coursework (Past and Current): Core PhD Micro and Econometrics Sequences, Industrial Organization (Theory and Empirical), Non-linear/Structural Econometrics, Statistical Inference, Advanced Natural Language Processing, Experimental Design, and High Dimensional Causal Inference (Causal ML)

MSc. in Computational Statistics and Machine Learning, Distinction — London, UK
University College London — Sep. 2016 – Sep. 2017

- Relevant Coursework: Supervised Learning, Graphical Models, Applied Machine Learning, Intro to Deep and Reinforcement Learning, Numerical Optimization, and Applied Bayesian Statistics.

B.A. in Economics, Cum Laude — New York, NY
Columbia College, Columbia University — Sep. 2011 – May 2015

WORKING PAPERS

- COVID-19 and remote work: An early look at US data. *No. w27344. National Bureau of Economic Research, 2020*
- Experiments in Algorithmic Collusion and Deep Reinforcement Learning
- Firm2Vec: Measuring Firm Differentiation with Industry Embeddings

PROFESSIONAL EXPERIENCE

Analyst — Aug. 2015 – Jul. 2016
Cornerstone Research — New York, NY

- Provided financial and economic analysis for commercial litigation matters.

Summer Analyst — Jun. 2014 – Aug.2014
JP Morgan Corporate & Investment Bank — New York, NY

- Global Trade and Loan Products Division

RESEARCH EXPERIENCE

Pre-Doctoral Fellow — Jul. 2017 – Jul. 2019
Stanford Institute for Economic Policy Research — New York, NY

- Provided research assistance to joint project by David Chan and Maria Polyakova.
- Research project: empirical investigation of Medicare quantitative regulation and its effects on physician behavior.

SKILLS AND OTHER

CS Competencies: Python (PyTorch, Keras), R, STATA, LaTeX, Julia, Linux.
Foreign Languages: English (Native), Spanish (Native), Mandarin Chinese (Native), French (Elementary).
Citizenship: Spanish.

REFERENCES

Doctoral Faculty Advisor
Prof. John Horton
MIT
✉ jjhorton@mit.edu

External Doctoral Faculty Advisor
Prof. Erik Brynjolfsson
Stanford University
✉ ErikB@stanford.edu