Data Analytics on Taxi Fare Prediction / Optimization

Study Ride-on-Demand Fare Algorithm to understand how a dynamic price is determined at the optimized level

Avinash Alapati, Bhargava Gaggainpali, Jubyung Ha, Pradeep Jaladi, Vilius Smilinskas

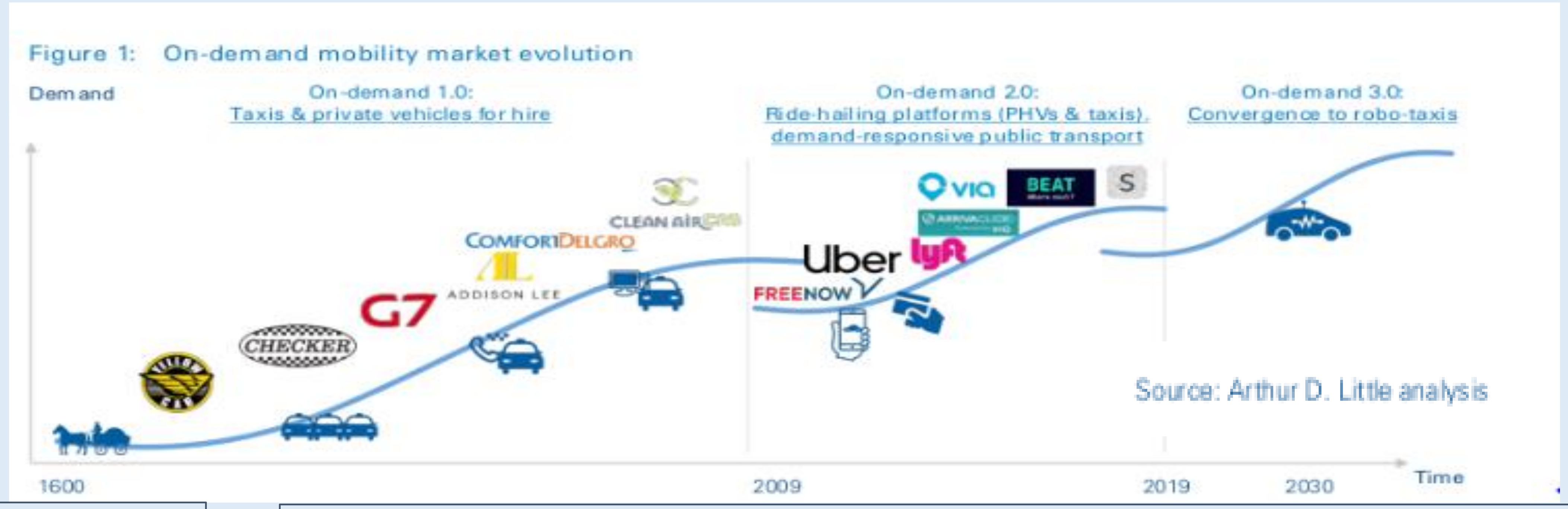
Introduction

Uber and Lyft are the best-known ride-on-demand (RoD) service companies compete in many different cities in various countries against each other and also against other local RoD companies. While passengers and travelers have many types of transportation services including subways, buses, taxies, and also RoD service companies, the reason that passengers and travelers use one of the RoD companies may stem from their edge-cutting convenience but also competitive price. When users open the Uber or Lyft apps and check the prices, they will realize the prices are constantly updated based on many factors. Even if all conditions are same, they could hardly find the same prices. In order to increase market shares and make profits; RoD Service companies would need to implement a sophisticated pricing program where they can gain competitive advantages on fares over other companies. Prices can be determined by many different attributes, such as car size, car model, a history of driver, distances, locations, special events, and so on. In this paper, we will conduct a research How each RoD company implements data science techniques to determine dynamic prices at the optimized level, meet customer demands, and increase their market shares?

Why is this data science and can data science help?

In the current 'Data' world, prediction is all based on the analysis of data available and collected. Taxi company's like Uber & Lyft, heavily depend on the data which is generated by their drivers to provide reliable service. Starting with general data like Pickup and destination is collected when the ride is booked and the demand is calculated based on the patterns (day, time, events..). Taxi drivers generate data even when they are not carrying any passengers like inferences on traffic patterns.

Data Science helps in prediction of supply and demand algorithm analysis, surge pricing, better cars, estimating fares, driver ratings and importantly for autonomous car research, tracking the location of drivers, monitoring driver's speed, motion and acceleration and identifying if a driver is working for a competing cab sharing company.



Data Understanding

On-Demand Rider Service needs to Understand the Underlying Data and services Required for Development, some of the important Data generating Components which needs to be taken in consideration are

- Map, Routing, GPS Tracking: Incorporating GPS in to On-Demand Ride Service for Routing and GPS tracking, Route based information for a user is stored in database.
- Push notification: Rider books a Ride a notification is sent to Driver, Payment: Payment made from Rider to Driver through Payment services,
- Driver's profile: Keeping track of Drivers Profile
- Ride Scheduling: Ride Scheduling component allows to schedule rides well ahead giving the riders advantage to riders to avoid price surge, User preferred timings and routes are stored in a database for future predictions.

Data Preparation

On Demand Rider has a massive database of drivers, so as soon as you request a car, On Demand Rider Algorithm matches you with the driver closest to you. In the background On-Demand Rider App is storing data for every trip taken — even when the driver has no passengers.

- Data is stored and leveraged to predict supply and demand, as well as setting fares.
- Rider Service also looks at how transportation is handled across cities and tries to adjust for bottlenecks and other common issues.
- Rider Service also gathers data on its drivers. In addition to collecting non-identifiable information about their vehicle and their location.
- Rider Service also checks for speed and checks to see if they are working for a competing company as well.

Evaluation

The work and the steps mentioned prior have to all be considered and monitored for proper results to be expected. Then the information can be processed and iterated upon to look for optimizations and solutions.To understand whether the provided solutions were effective, the RoD provider needs to have metrics and expectations set out for the work to be evaluated. The RoD provider needs to monitor its Data Streams to determine customer satisfaction and performance efficiencies. The major monitored aspects are:

- Customer Satisfaction
- Supply and Demand
- Ride Fares
- Route Efficiency

Deployment

Combining the information from these multiple sources means that the work that is finished and evaluated can be comprehensive and utilized in for many different solutions. Most importantly the deployment of the solution should be in real time data processing. Algorithms should be updated and set up after diligent testing and evaluation of performances. Due to utilizations of these algorithms RoD providers are able to uniquely position themselves within the market. Due to readily accessible demand from customers, being efficiently routed to drivers, that then take the optimized route with fares that are beneficial for both the customer and the provider.

Data Modeling

On-Demand Rider is built on a machine learning platform as well as natural language and dialog system technologies.

- ML Algorithm automatically processes and verifies millions of business-critical images and documents such as drivers' licenses and restaurant menus, among other items, per year.
- ML Algorithm to be used for sensor processing, crash detection and safety optimization.
- Algorithms match drivers with passengers based upon their location, needs and availability
- Routing technologies are also based on algorithms that handle thousands of ETA requests per second.
- Types of predictive models and probability creations include Destination Prediction, Michelangelo, and the Bayesian neural network (BNN) architecture.

Conclusion

Every day, millions of passengers travel around in their city, but no two identical riders share the same experience as others. Developing an ML Algorithm which can predict the rides effectively will help both drivers and customers. Some of the benefits are as follows.

Dynamic Pricing: Increase in Price is often predicted on different parameters like weather conditions, Traffic Patterns, Time of the day.

If there is a price surge, the no of drivers will move towards hot spots, only the customers with great need will prefer to travel while other customers will wait for a bit. If the available riders started to increase in the hot spot area, the prices might come down a bit for the customers who are waiting for better pricing.

Upfront pricing: Based on the past historical data available, if a customer can see the price of ride, the Customer can plan his accordingly as per prices well in-advance; this is known as Upfront Pricing.

Route-based pricing: If a customer takes a ride to a city center, there is a greater possibility of pooling the ride with other customers. The prices might come down also the next immediate ride is likely to occur at the drop off point. But Customer takes rides further away to the countryside the chances of next immediate as slim.

Acknowledgements

We thank all the authors who wrote and provided helpful comments on previous versions of this document. All authors did an extended analysis on the data which is available in the market to find out how the taxi fare calculation is done by Uber and Lyft to make more profits. Some of the references cited in this paper are included for illustrative purposes only.

References

1. Hussan Kazmi US20150348332A1 - Method and apparatus for taxi cab fare collection. Retrieved January 5, 2020 from https://patents.google.com/patent/US20150348332A1/en
2. Pavlos Protopapas Kevin Rader Margo Levine Rahul Dave Introduction to Regression Data Science https://harvard-iacs.github.io/2017-CS109A/lectures/lecture4/presentation/lecture4.pdf
3. Ying-Wen Bai and En-Wen Wang. 2012. Design of taxi routing and fare estimation program with re-prediction methods for a smart phone. 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings(2012). DOI: http://dx.doi.org/10.1109/i2mtc.2012.6229165
4. Nikita Shahane, Megha Kaul, and Yong Zheng. 2019. Exploratory Analysis of Chicago Taxi Rides. Proceedings of the 20th Annual SIG Conference on Information Technology Education - SIGITE 19(2019). DOI: http://dx.doi.org/10.1145/3349266.3351390
5. Jintao Ke, Hongyu Zheng, Hai Yang, and Xiqun (Michael) Chen. 2017. Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. Transportation Research Part C: Emerging Technologies85 (2017), 591–608. DOI: http://dx.doi.org/10.1016/j.trc.2017.10.016
6. Wenbo Zhang, Dheeraj Kumar, and Satish V. Ukkusuri. 2017. Exploring the dynamics of surge pricing in mobility-on-demand taxi services. 2017 IEEE International Conference on Big Data (Big Data)(2017). DOI: http://dx.doi.org/10.1109/bigdata.2017.8258070
7. Peter Cohen, Robert Hahn, Jonathan Hall, Steven Levitt, and Robert Metcalfe. 2016. Using Big Data to Estimate Consumer Surplus: The Case of Uber. (2016). DOI: http://dx.doi.org/10.3386/w22627
8. Ce Liu and Qiang Qu. 2015. Trip Fare Estimation Study from Taxi Routing Behaviors and Localizing Traces. 2015 IEEE International Conference on Data Mining Workshop (ICDMW)(2015). DOI: http://dx.doi.org/10.1109/icdmw.2015.106
9. Wei Zhang, Ruichun He, Yong Chen, Mingxia Gao, and Changxi Ma. 2019. Research on Taxi Pricing Model and Optimization for Carpooling Detour Problem. Journal of Advanced Transportation2019 (2019), 1–11. DOI: http://dx.doi.org/10.1155/2019/3867874
10. Yuandong Wang et al.2019. A Unified Framework with Multi-source Data for Predicting Passenger Demands of Ride Services. ACM Transactions on Knowledge Discovery from Data13, 6 (2019), 1–24. DOI: http://dx.doi.org/10.1145/3355563
11. Christophoros Antoniades, Delara Fadavi, Antoine Foba Amon Jr. 2016. Fare and Duration Prediction: A Study of New York City Taxi Rides. DOI: https://pdfs.semanticscholar.org/6b2b/99187100b097978f4659c1b655b7c932c7de.pdf
12. Hyoung Suk Shim. 2017. Taxi Fare Mechanism: New York City Yellow Cab Trip. DOI: https://pdfs.semanticscholar.org/222f/ec4e1c2911de935dd4afacff6be51b55eacb.pdf
13. Baozhen Yao, Lu Jin, Qingda Cao, Junjie Gao and Mingheng Zhang. 2016. Fleet size and fare optimization for taxi under dynamic demand. DOI: http://www.scielo.br/scielo.php?pid=S2238-10312016000400045&script=sci_arttext
14. Gyozo Gidofalvi1 and Torben Bach Pedersen. 2007. Cab-sharing: An Effective, Door-to-Door, On-Demand Transportation Service. DOI: http://www.diva-portal.org/smash/record.jsf?pid=diva2%3A500467&dswid=-9354
15. Scott Wallsten. 2015. The Competitive Effects of the Sharing Economy: How is Uber Changing Taxis?. DOI: https://techpolicyinstitute.org/wp-content/uploads/2015/06/the-competitive-effects-of-the-2007713.pdf
16. Bhaskar Basu. 2017.UBER: Redefining Transportation. DOI: https://www.academia.edu/download/54901145/Dr._Bhaskar_Basu_-_V5I3-0013.pdf