From Data Science to Blockchain - Analytics in Cross-Border Logistics

Zhuojie Huang
Commerce Cloud Data
Pitney Bowes Inc.
Shelton CT United States
zhuojie.huang@pb.com

ABSTRACT

Cross-border ecommerce has expanded at an exceptional rate. It is projected that global shipping volume will reach 100 billion by 2020. Nowadays, millions of customized, small, individual packed parcels flow through a complicated logistic supply chain involving local, ground and air carriers from their original countries to the consumers in another destination country. Accomplished by the physical flow, billions of data points for these parcels move across countries digitally. This research review the end to end physical and digital cross-border logistics supply chain, and identify several data science use cases for dynamic routing/pricing, demand forecasting and parcel flow prediction. Current logistic entities in this supply chain has a limitation on data visibility and transparency, as a result, they are restrained to solve a local problem under their regime. It is envisioned that the mature of blockchain technology will disrupt the traditional linear logistics supply chain by providing distributed verified information in real time. Different participants in the cross-border supply chain will benefit from the consensus information. In the future, data science models in this field will evolve and focus more on real time analytics and aiming towards a global optimization.

CCS CONCEPTS

• Applied computing → Electronic data interchange • Applied computing → Transportation

KEYWORDS

Data Science, Operation Research, Cross-border Logistics, Blockchain

ACM Reference format:

Zhuojie Huang. 2019. From Data Science to Block Chain - analytics application in cross-border logistics. In *Proceedings of the 8th SIGKDD International Workshop on Urban Computing (UrbComp'19). ACM, Anchorage,Alaska, USA, 4 pages.*

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
UrbComp'19, August, 2019, Anchorage, Alaska USA
© 2019 Copyright held by the owner/author(s)

1 Introduction

Cross-border ecommerce grows at an unprecedented pace, as it is projected that global shipping volume will reach 100 billion by 2020 [1]. Various flourishing ecommerce platforms enable consumers easily order merchandise from manufacturers, immediate distributor or shops from its originated country at any time. Every second, 2300 parcels are shipped globally[1]. These small, individual packed parcels, usually confined with a delivery time windows, go through a complex international trade network from the source manufacturer to the end consumers in another country. Along with the physical flow, billions of data points pour directly into the downstream shippers and carriers so they can react, plan their resources, and ensure the parcel meeting the service level agreement. Afterward, a cash flow will be initiated across country as the fulfillment of the contract. Since most of the transaction happen at the real time, traditional linear, isolated and disconnected supply chain has to be transformed through digital interconnected devices and complex networks to enable the complex information exchange [2]. With abundant logistics data, data science plays a more and more role in cross-border logistics, enabling merchants, shippers and carriers for network analysis, resource allocation and optimization, risk prediction, demand forecasting and flow predictions [3].

On the other hand, getting accurate and verified supply chain data remains a challenge [4]. The blockchain technology, first matured at the field of cryptocurrency currency, empowers a distributed, digital ledger to keep track for any data exchange, agreements/contracts, tracking and payment [5]. The application of block chain can orchestrate and synchronize physical flow of goods with information and financial flows, and help to alleviate the data quality challenge and enhance trust.

This research review the end-to-end cross-border supply chain and identify several use cases for the data science applications in the cross-border logistics. Then I will discuss the emergence of blockchain technology and present ideas and suggestions for the future research directions.

1.1 Parcel movements on physical and digital network

The physical movement of a parcel can be represented by a graph G(V,E) where V is a set of locations that can interact with parcels and E represents the method for a parcel that move from one node to another node. Inside a node V, a subnetwork represents the workflow associated with freight unloading, sorting, staging, containerization and freight loading. This digital representation of the network have two scales, and requires the combination of the sensor data from GPS and in node data from Internet of Thing devices and RFID to gather the real time information on the location and the status of the parcel.

Figure 1: A parcel journey from China to United State

Figure 1 shows a parcel journey from the manufacturer from Yi Wu, China to the end consumer in Philadelphia, United States. The parcel was picked up by a local carrier in China and sent to Shanghai by a truck. Then the parcel was inducted by an air freight-forwarder, went through the China customs clearing and flew to Taipei for consolidation. After it stopped at Anchorage Alaska, it continued its journey and arrived at John Kennedy Airport at New York City. Afterward it was electronically reviewed by US customs and collected by another carrier before its final delivery to Philadelphia. At each location, a process inside a facility was initiated to sort and touch the parcel to ensure it is on a predefined route to its final destination.

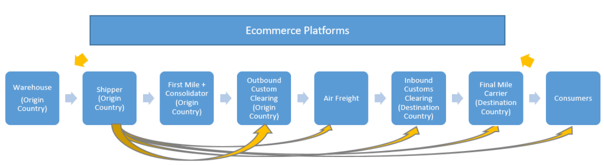

Figure 2 Physical and digital flow for cross-border ecommerce. Blue represents the physical flow and yellow represents an example for the data flow from one party to downstream

Accomplished with the physical movement from depot to depot, virtual properties of the parcel move on a digital network at a faster speed. The digital twin representation emphasize the visibility and transparent of all the information for all stakeholders including local carriers, freight forwarder, customs and consumers [4]. The blue arrows in Figure 2 show a linear example for the physical from the origin country to consumers in another country. Ideally, entities in this chart should be informed about parcel's real time location and perform their function properly.

The yellow arrows in Figure 2 shows the data flow from the upstream shipper to multiple entities downstream. With mutual benefits, shippers at the upstream have incentives to share parcel information as it will improve the overall supply chain efficiencies. With this information on hand, downstream entities (carriers/shippers) can improve their operational efficiencies by unifying production optimization, scheduling high productivity work shift, maximizing utilization of their facility space, coordinating truck inbound and out bound, and optimizing inter node logistics planning [6]. Air freight forwarders can optimize the price to order the space on airline and maximize the container utilization with weights, current capacities and air compliance constraints information [7]. Customs and Border Protection (CBP) can utilize the information to facilitate import entry and identify the possibilities for violation to ensure speedy custom clearance. At the end of the chain, consumers can obtain their parcels' real time locations, level set their expectation on parcel arrival time and improve overall customer experiences.

Although accurate, timeless, consistent and completeness data are expected, missing data along the parcel journey is inevitable as the parcel is handed over multiple times to different entities and interacting with different information system – a lot of times this process is still handled manually. Recent advancement in Recurring Neural Network [8] shed some light on improving multivariate time series prediction accuracy with the imputation of missing data. As identified in [4] that missing data remain a consistent challenge, this requires that stakeholders to make managerial decision with the adaption of a more robust data analytic-based strategies to tackle this challenge.

1.2 Dynamic routing and dynamic pricing

During the cross-border transportation process, every individual parcel has it total time restrictions: the length of any route may not exceed a prescribed bound L; this length consists of intercity/depot travel times cij and of processing time ti at each city i on the route. Some depots, such as customs export and import cannot be bypassed, as they are mandatory requirements for the parcel delivery. Since the participated entities only own part of the supply chain, they are limited to perform local optimization –given the current position of the parcel, which route should be chosen next to meet the total time restriction to the consumer. Often, they need to deal with network disruption such as hazardous weather event and service down time by changing a route. Understand how the parcel can be dynamically route at each touch point, and dynamically choose the next service provider is crucial for the logistics entities here, and remains a challenge.

As the parcel move from one carrier to another, they are following a predetermined contractual rates. Given the fluctuation of the supply and demand, there are lot of uncertainty of profitability for the shipper and raising the cost for a consumer. Based on a preset of the origin and destination, for each hop, carrier/shippers should be able to look for available capacity, and determine the next carrier based on the current and historic service level agreement, speed and service reliability. Research has utilized deep learning models for dynamic pricing [9]. It is foreseeable that a real time bidding system to balance supply and demand as similar in the advertisement industry will be implemented in the cross border logistics service. The bidding mechanism can lower the cost for shipping and improve the overall supply chain efficiencies.

1.3 Forecasting demand patterns and predicting parcel flow

Demand forecasting plays a central role for almost all supply chain related decisions, particular for logistics entities to facilitate resource planning, capacities utilization and local route optimization. Inside a warehouse, having an accurate prediction of the incoming parcel volume is critical for operational managers to plan staff properly, minimize cost and maximize resources utilization. From the technical perspective, recent application using Long Short Term Memory(LSTM) model shows promising results to include the spatial and temporal features to improve prediction accuracies[10]. Various deep learning models have been applied to predict ride share passenger demands[9], [11], [12]. However, the bullwhip effect, which describe that the demand variability as one moves up a supply chain has yet fully studied in this complicated cross border logistics environment. As discussed in [13], sharing demands information from the downstream entity to the upper entity can alleviate the problem. At the downstream of the supply chain, Ecommerce demands by consumers are spatially clustered and correlated to population distribution, and has time series components linked to brand affinity, seasonality, promotion and holiday schedule. How to account for the downstream variation and model the bullwhip effect for upstream entities to reduce demand uncertainties across logistics entities from different countries has yet fully studied.

Estimating the arrival time for the each hop and detect potential parcel journey abnormality has immense practical value in supply chain planning. Prediction the next stop for a trajectory is a well-studied area in GIS literatures [14], [15]. Various deep-learning framework have been proposed to add spatial/temporal information for time-series forecasting [16]–[18]. Though less uncertainties in path determination in the cross border logistic network, the current deep learning for time series prediction framework can still be applied to the supply chain analysis and identify the abnormal parcel and potential path derivations.

1.4 Block chain, consensus and compliance

In the cross-border logistic network, parcels and information travel from one entity to another entity and from one country to another country. Thus, it requires different levels of verification to ensure the fulfilment of the contracts with traceable chain of custody : 1) Compliance verification of import/export requirements. CBP needs to ensure that product is compiled to local export/import law. 2) Parcel attribute verification: when a parcel's custody has been handed over from one entity to another, weight, dimension need to be verified by different carrier and to minimize discrepancies. 3) Location verification of the parcel to match physical and digital property and to ensure the parcel is on route and can be delivered under its current service level agreement. 4) Transaction verification to make sure the invoicing is correct and the cash flow can seamlessly move from one country to another country with a consent exchange rate.

Blockchain can ensure trust and improve data integrity amongst different entities in the cross-border logistics network. As discussed in [19], blockchain can simplify the global trade, improve transparencies and tractability in supply chains, and automated process for smart contract. The traceability of the data will further enhance model interpretation and generate useful insights for business decision for all the logistics entities from either upstream/downstream of the supply chain from different countries. More importantly, these information will be presented and updated in real time so every entities have most up-to-date consensus, verified information to optimize their operations and improve their network efficiencies.

One of the proposed use case for using blockchain is to facilitate the custom clearing process when a parcel is moved from one country to another country. The blockchain application can enable the sharing of merchant, product and shipment information with verified name, address, contact information, business type and other agreed upon information amongst CBP and different shipping entities. As the physical shipment started, information for the parcel can be place on the on the blockchain with clear indication of source, item description, item category and the tracking information of its chain-of-custody. Through the blockchain, CBP would have the ability to provide feedback to shippers and merchants related to missing or incorrect information or provide other instructions for custom clearance. While the merchants and shippers make changes and update additional information, this would also be captured on the blockchain and other entities will be notified for the updates. As mutual benefits, more accurate data with verified information are shared between CBP, merchants and shippers with less errors, which will reduce number of holds for inspection and the time to clear a parcel.

2 Discussion

Today, most of cross-border logistics entities in the supply chain act as isolated blocks while lacking a centralized way to share trusted information. Applying the optimization model and predictive models within these entities can provide a local solution to solve one specific problem in the chain. Backed by the blockchain technology, a more connected supply chain will soon disrupt the traditional linear model. Combined with data from Internet of Thing (IoT), GPS and robotics information, block chain technology provides a viable option to orchestrate and synchronize the physical flow of the good with data flow and financial flow. Ultimately, all the entities in the cross-border supply chain will consolidate order data, location data, in node sensory data, financial contract data and transaction data in a distributed manner with agreement on levels of details to share. Data scientist can leverage this real time, better-integrated and traceable data for real time analytics and optimization, thus aiming towards a global optimization model.

ACKNOWLEDGMENTS

I would like thanks Rick Ryan for his comments and suggestions on the blockchain use case. I would also thank Brian Lemm for his review of this paper. Last, I would also thank Olga Lagunova, Surya Sagi and David Messineo for the support of this research.

REFERENCES

[1] Pitney Bowes, “Pitney Bowes Parcel Shipping Index.” [Online]. Available: https://www.pitneybowes.com/us/shipping-index.html. [Accessed: 10-May-2019].

[2] R. De Souza, L. William, and G. Timperio, “Supply Chain Digital Transformation: Insights and Tools – anyLogistix Supply Chain Optimization Software.”

[3] M. A. Waller and S. E. Fawcett, “Data science, predictive analytics, and big data: A revolution that will transform supply chain design and management,” *J. Bus. Logist.*, vol. 34, no. 2, pp. 77–84, Jun. 2013.

[4] B. T. Hazen, C. A. Boone, J. D. Ezell, and L. A. Jones-Farmer, “Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications,” *Int. J. Prod. Econ.*, vol. 154, pp. 72–80, Aug. 2014.

[5] Bernard Marr, “How Blockchain Will Transform The Supply Chain And Logistics Industry,” *Forbes*, 2018. [Online]. Available: https://www.forbes.com/sites/bernardmarr/2018/03/23/how-blockchain-will-transform-the-supply-chain-and-logistics-industry/#3b2ff7e95fec. [Accessed: 12-May-2019].

[6] R. Y. Zhong, G. Q. Huang, S. Lan, Q. Y. Dai, C. Xu, and T. Zhang, “A big data approach for logistics trajectory discovery from RFID-enabled production data,” 2015.

[7] M. Mongeau and C. Bes, “Optimization of aircraft container loading,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 1, pp. 140–150, Jan. 2003.

[8] Z. Che, S. Purushotham, K. Cho, D. Sontag, and Y. Liu, “Recurrent Neural Networks for Multivariate Time Series with Missing Values,” *Sci. Rep.*, vol. 8, no. 1, p. 6085, Dec. 2018.

[9] M. F. Dixon, N. G. Polson, and V. O. Sokolov, “Deep learning for spatio-temporal modeling: Dynamic traffic flows and high frequency trading,” *Appl. Stoch. Model. Bus. Ind.*, Sep. 2018.

[10] F. Karim, S. Majumdar, H. Darabi, and S. Chen, “LSTM Fully Convolutional Networks for Time Series Classification,” *IEEE Access*, vol. 6, pp. 1662–1669, 2018.

[11] L. Moreira-Matias, J. Gama, M. Ferreira, J. Mendes-Moreira, and L. Damas, “Predicting taxi-passenger demand using streaming data,” *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 3, pp. 1393–1402, Sep. 2013.

[12] J. Ke, H. Zheng, H. Yang, and X. (Michael) Chen, “Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach,” *Transp. Res. Part C Emerg. Technol.*, vol. 85, pp. 591–608, Dec. 2017.

[13] F. Chen, Z. Drezner, J. K. Ryan, and D. Simchi-Levi, “Management Science Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information,” 2000.

[14] Q. Huang, “Mining online footprints to predict user’s next location,” *Int. J. Geogr. Inf. Sci.*, vol. 31, no. 3, pp. 523–541, Mar. 2017.

[15] J. J.-C. Ying, W.-C. Lee, T.-C. Weng, and V. S. Tseng, “Semantic trajectory mining for location prediction,” *Proc. 19th ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst. - GIS ’11*, p. 34, 2011.

[16] J. Zhang, Y. Zheng, D. Qi, R. Li, X. Yi, and T. Li, “Predicting citywide crowd flows using deep spatio-temporal residual networks,” *Artif. Intell.*, vol. 259, pp. 147–166, Jun. 2018.

[17] Z. Pan, Y. Liang, J. Zhang, X. Yi, Y. Yu, and Y. Zheng, “HyperST-Net: Hypernetworks for Spatio-Temporal Forecasting,” Sep. 2018.

[18] J. Zhang, Y. Zheng, J. Sun, and D. Qi, “Flow Prediction in Spatio-Temporal Networks Based on Multitask Deep Learning,” *IEEE Trans. Knowl. Data Eng.*, pp. 1–1, 2019.

[19] DHL, “Blockchain in logistics,” *DHL Cust. Solut. Innov.*, pp. 1–28, 2018.