

# Harnessing Big Data to Identify Ideal Locations for Warehouses and Bike Share Stations

Dr Hui Xiong





# HARNESSING BIG DATA TO IDENTIFY IDEAL LOCATIONS FOR WAREHOUSES AND BIKE SHARE STATIONS

With a sharp increase in the public use of online shopping in recent years, which has spiked further due to the COVID-19 pandemic, the importance of warehouse positioning to provide an optimal delivery service has become a significant area of focus for retailers. Similarly, bike sharing in major cities has also seen an astronomical rise in usage – prompting questions about how bike stations can be best positioned. Stations in optimised locations would require minimal interference to change the stock levels, while also ensuring the greatest accessibility to users. **Dr Xiong** and his colleagues at Rutgers University have been using the ubiquity of big data to better optimise these systems for both cost efficiency and standard of service to customers, testing their findings in real-world scenarios.

---



## Warehouse Site Selection for E-Commerce

Being an extremely competitive market, e-commerce businesses must find slight advantages over competition in order to boost their profits. One of the biggest draws for customers who shop online is fast delivery – the convenience of ordering an item one day and having it delivered the next.

To make this a viable option, e-commerce companies have to be able to effectively tackle the so-called

‘connected capacitated warehouse location problem’. This, in its essence, aims to minimise the costs of shipping (supplier to warehouse), delivery (warehouse to customer) and the inter-warehouse transportation involved in keeping stock levels adequate. Optimal warehouse positioning is pivotal in making this happen.

A big first step in choosing warehouse locations is being able to predict how sales will be distributed geographically, depending on how effectively retail delivery can operate. However, most

available predictive models struggle to find patterns between different logistical factors and customer satisfaction, and so are strictly limited in their prediction accuracy and usability.

## A New Model for Warehouse Locations

In a paper for the 2017 International Conference on Data Mining, Dr Xiong and his colleagues at Rutgers University provide a ‘data smart approach’ to the connected capacitated warehouse location problem. Firstly, the researchers address the problem of predicting customer demand. ‘When all other factors are equal – including retailer reputation score and product quality evaluation – customers are more willing to choose the retailers with fast shipping, fast delivery and high-quality logistics,’ Dr Xiong notes.

As such, his team has built their predictive model around three factors that are known to influence customer demand: the time taken for an item



to be shipped after an order is confirmed, the time between an order being shipped and arriving at a customer, and the prevalence of damaged items after being shipped. The model is then driven by an artificial neural network – a network inspired by the human brain that allows computers to learn patterns and identify relationships based on input data, and subsequently carry out a human-like decision-making process. For Dr Xiong's research, this means using real-world logistics data for the three influencing factors previously listed and then producing a prediction of customer demand.

In fact, during their study, Dr Xiong's team put their new model to the test, using data from almost 3.5 million transactions that took place in 2012 on a Chinese online shopping site called Xiaoye. Using other predictive algorithms for comparison, the researchers showed that their neural network model was significantly more accurate, with an error rate of 27.25% – compared to 28.99% achieved using the next-best model.

Interestingly, their analysis also showed that customers are more likely to opt for a delivery service with shorter shipping time but longer delivery time than the other way around. This indicates that the transport of goods from supplier to warehouse is of greater importance than that from the warehouse to customer, meaning that keeping warehouse stock levels consistent is paramount.

This finding is the main driver in the team's algorithm for choosing optimal warehouse locations. Their algorithm keeps computational costs about 40% lower than alternatives, and works on the principle of assigning customers to warehouse locations that minimise logistics costs. Ultimately, Dr Xiong and his colleagues showed that having a network of four connected warehouses is much more effective in managing stock levels and keeping shipping costs low than having one centralised warehouse – bolstering the argument for warehouse networks.

### **The Rise of Bike Share Systems**

Another important application of the research carried out by Dr Xiong and his colleagues is bike sharing. Bike sharing systems are now prevalent in many major cities around the world, with



their usage in New York alone accounting for 17.58 million trips across 329 stations between the summers of 2013 and 2015. They appeal to both commuters and tourists alike, providing a healthy and cheap option for short distance journeys and filling in gaps where public transport is limited or unavailable. Just like the surge in online shopping, the use of bike sharing systems has also witnessed a significant rise during the COVID-19 pandemic, as they provide a far safer alternative to public transport.

As a result of their increasing popularity, many cities such as New York are looking to expand their current system and implement more pickup and drop-off locations for the bikes. However, this comes with numerous challenges.

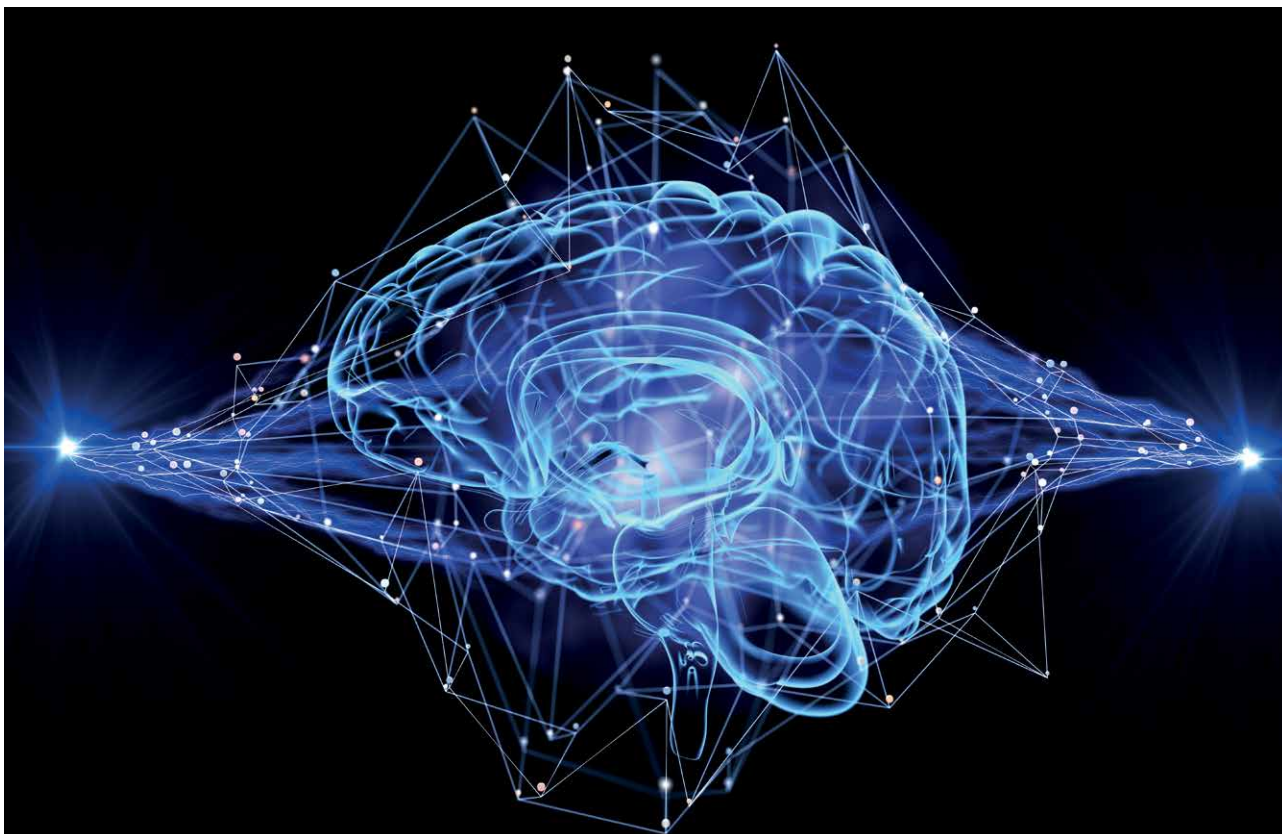
The new stations need to be effective in providing the public with access to bikes in locations that are prone to high demand, such as those between subway stations and corporate buildings for morning commuters. At the same time, the stations need to maintain an acceptable stock level – ideally never short of bikes and also never becoming overfilled. In other words, new stations should allow for a self-sustaining network where incoming bikes replace outgoing ones. If this is not the case, rebalancing operations must take place, which involves transporting bikes from one location to another – an additional expense for companies running the bike stations. This system should of course come hand in hand with the bikes being available at busy locations and covering popular journeys.

Taking this into consideration, bike share companies need to use a predictive model that can take into account a variety of influencing factors and guide the positioning of stations within cities. So far, the current models have only been able to use bike usage patterns and station demand data, whereas influencing factors of geographical bike demand have not been properly investigated.

### **Rebalancing Bike Share Systems**

Using a similar artificial neural network model to that used for warehouse positioning, Dr Xiong's team has also produced





remarkable results in predicting customer demand for bike sharing systems, and keeping stock levels balanced at stations. This time, the data inputted into the artificial neural network included walking distances between bike stations, walking distances from bike stations to underground entrances, density of taxis in particular regions and points of interest in particular regions.

Applying the model to data from 256 of New York's CitiBike stations, the team showed that their neural network model had an accuracy of 85.2% in predicting bike demand, revealing 185 stations as being balanced and 71 as unbalanced. The level of accuracy achieved was a significant improvement over alternative models.

Showing further advancements in this area, Dr Xiong and his colleagues have played an integral role in solving the rebalancing problem – that is, minimising the costs in rebalancing bike stock – by implementing a new programming method.

As noted by Dr Xiong, 'Clustering can be used to reduce the complexity of the large-scale optimisation problem.' Clustering works by grouping certain bike stations together, and having the stock in these particular groups serviced by one vehicle. The stations are grouped based on functional zones – areas that have similar surrounding points of interest and are located relatively close-by to one another.

Certain functional zones will contain self-sustaining stations that don't need to have their bike stock rebalanced by non-users. Others will need vehicles to deliver or collect stock to maintain appropriate numbers of bikes. This method effectively changes a large-scale problem of great complexity with bike stock controlled by multiple vehicles, into simpler problems for individual areas with stock controlled by only one vehicle. Once this is achieved, the vehicle rebalancing route can be optimised to be significantly more efficient.

Using real-world bike trip data and weather data covering 328 bike stations in New York's CitiBike scheme, the new vehicle routing method – based on clustering – proved to be much more effective than the so-called Nearest Neighbour Insertion Algorithm – a routing method generally praised for being an effective solution to the problem. This is thought to be due to the small distances covered in the clustering scheme compared to the long-distance nature of the Nearest Neighbour Insertion Algorithm, which adds complexity and consequently cost to the problem.

### Into the Future

Through their extensive research, Dr Xiong and his colleagues have demonstrated effective solutions to both the connected capacitated warehouse location problem and the bike share rebalancing problem. As the COVID-19 pandemic has caused an increased need for optimised bike sharing schemes and online shopping, the team's research comes at a pertinent time, and will likely lead to great advancements for bike hire and retail companies alike.



# Meet the researcher

**Dr Hui Xiong**

Rutgers Business School

Rutgers, the State University of New Jersey

New Jersey

USA

Dr Hui Xiong graduated with a PhD in Computer Science in 2005 from the University of Minnesota – Twin Cities. A recipient of the IEEE Fellow award and the Ram Charan Management Practice award from Harvard Business Review, among many other individual accolades, Dr Xiong received two-year early tenure in 2009 and the RBS Dean's Research Professorship in 2016 at Rutgers University. His research primarily focuses on the application of data analytics for developing efficient solutions for data intensive tasks. An esteemed publisher of numerous peer-reviewed and conference papers, Dr Xiong has also edited *Encyclopaedia of GIS*, a top 10 most popular computer science book written by Chinese Scholars as recognised by the major publisher Springer. In addition to his research, which garners national and international attention, Dr Xiong is heavily involved in mentoring PhD students, 16 of whom have graduated to date and the majority of whom have become tenure-track professors in renowned universities in the world.

## **CONTACT**

**E:** [hxiong@rutgers.edu](mailto:hxiong@rutgers.edu)

**W:** <http://datamining.rutgers.edu/>

## **KEY COLLABORATOR**

Dr Junming Liu, Assistant Professor, Department of Information Systems, City University of Hong Kong

## **FUNDING**

US National Science Foundation

## **FURTHER READING**

J Liu, L Sun, Q Li, J Ming, Y Liu, H Xiong, Functional Zone Based Hierarchical Demand Prediction for Bike System Expansion, KDD '17: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2017, 957–966. <https://doi.org/10.1145/3097983.3098180>

C Chen, J Liu, Q Li, Y Wang, H Xiong, S Wu, Warehouse Site Selection for Online Retailers in Inter-connected Warehouse Networks, 2017 IEEE International Conference on Data Mining, 805–810. <https://doi.org/10.1109/ICDM.2017.96>

J Liu, L Sun, W Chen, H Xiong, Rebalancing Bike Sharing Systems: A Multi-source Data Smart Optimization, KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2016, 1005–1014. <https://doi.org/10.1145/2939672.2939776>

J Liu, Q Li, M Qu, W Chen, J Yang, H Xiong, H Zhong, Y Fu, Station Site Optimization in Bike Sharing Systems, 2015 IEEE International Conference on Data Mining. <https://doi.org/10.1109/ICDM.2017.96>

