Application of Machine Learning and Real-time Feedback System to Predict Arriving Time

Kaiwen Guo
Walsh Jesuit High School, Hudson, Ohio
vivian.wang@cas-harbour.org
* Corresponding author: kevinguo2021@gmail.com

Abstract—Real-time location feedback provides customers with direct information about the location of the package, which also provides one of the factors needed to predict the arrival time. This is not a simple math problem once we have the real-time location data because using math equations could not account for different situations that might occur to packages, which would again provides an inaccurate delivery time. By using machine learning, based on the data collected earlier and the information about the location from the use of real-time location feedback system, computers would be able to predict a delivery time with different factors involved, including weather and the possibility of the accident happened on the way. Using those different skills and factors, customers would be able to get an arrival time that is more precise than before and able to do decisions based on that. The main idea of this article is to predict the delivery time with the combination of the real-time location feedback system and machine learning.

1. INTRODUCTION
Many people find difficulties when they bought products online but the sellers provide few data related to the delivery time. The absence of precisely predict delivery time affects many customers who decide to buy products at the last minute, such as buying products right before going on vacations. Most of the systems used in companies such as Amazon are not mature. The idea of their systems is that they want to provide a big number of dates so that customers would not have an unexpected time of delivery. They account every kind of situation that might affect the delivery time—rain, snow, crowding roads, and all kinds of other situations. But, for people who want to do a last-minute shopping, that method would not work well. As a matter of fact, most of the packages would arrive 4 to 5 days earlier than then “predicted delivery time” that the companies provide to customers. Some people would not purchase some products because they might think that the package would deliver after the need for it. They might lose the opportunity to enjoy the convenience of the product just because of the inaccurate delivery time. To that end, I see a strong motivation to courage the use of new predict systems about delivery time. There are two important methods that need to be involved to build the new predict system.

Two important methods that involved in achieving this goal is the use of real-time on a feedback system and machine learning model. Real-time feedback system, including real-time location feedback system, which is a system and method is disclosed for tracking in real-time the location of a person, animal, or another object in a community having a cable infrastructure[1], and real-time weather feedback system, to help the machine learning model to adjust the weight of each factor especially
weather factors. The use of machine learning is also important to build a success delivery time predict model. Machine learning is a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty[2].

Application of real-time feedback systems and machine learning on delivery time has not developed in the United States much. Companies such as Amazon would only give customers a predicted time using data from the place of delivery to the place of receipt. They also account for any possible situations happening to the package, which involves with many low possibility existing situations and will make the delivery time unreliable. The delivery time might be four to five days longer than the real-time taken by the package to deliver. Not like companies in the United States who have not developed many skills in predicting the delivery time, companies in other countries like Taobao, a company in China, have developed a mature system in predicting the delivery time.

2. METHODOLOGY
In this predict system, the main idea is to predict the delivery time with the combination of the real-time location feedback system and machine learning.

Before the package starts shipping, the predicted system will first have an “overall” delivery time predicted. This is still based on the machine learning process, but this is trying to predict an overall delivery time, so still, it needs to account for all the possible situations happening to products while delivering. The best machine learning algorithm to use here is a decision tree, a hierarchical model for supervising learning whereby the local region is identified in a sequence of recursive splits in a smaller number of steps[3]. By using a decision tree, all the situations will each be part of the factors of the system and be listed from the highest possibility of existing to the least. Under each factor, the delay time is also listed to classify how long the delay is if this situation really occurred on the package. At this point, the delivery time predict system is trying to find an “overall” delivery time that includes all the possible situations, so the delivery time predict system need to pick the branch that all high existing possibility situations occurred, which has all situations have a higher possibility of existing than five to ten percent. The reason why the delivery time predict system don’t pick the branch that has all situations occurring is that it will provide us an extremely high number which is not helpful for customers to use to decide whether buying or not. This delivery time is inaccurate because there is a little possibility that all those situations accounted before exist, and other low existing possibility situations would also exist, so we need to fix the predicted delivery time as we move on.

![Figure 1. A decision tree classifier.][4]

Each box is a node at which tests (T) are applied to recursively split the data into successively smaller groups. The labels (A, B, C) at each leaf node refer to the class label assigned to each observation.

Now it is the time that real-time feedback systems take account. Every five to six hours, the real-time location feedback system will record the recent location of the package and send the data to
the predicted system. This data is important because previous data of delivery time from location to location is most likely different. And since the package already moves from the beginning location to another, it needs to predict the delivery time again using location information from that point. Even though the delivery time predict system is using a real-time location feedback system to record the location of the package once awhile, it is not going to record the specific point of the location at that time but a bigger area of the location. The reason for doing so is that it will be much harder to collect data if the delivery time system only record the specific point as the location factor. When zoom out to a bigger area to record, it would be easier for the machine learning model to study and be able to have a smaller dataset than a model that has record location more precisely.

But, not only real-time location feedback systems, other real-time feedback systems such as weather reports will also be taken account of. Real-time feedback Systems like weather reports take another big part of the methods. According to Edward N. Lorenz, weather’s predictability results because they are governed by physical laws which presumably do not change with time. These laws tell us that the past and future weather are in some way related”[5]. Then, the system will use machine learning, which has all the data from that point to the transfer station closets to the receiver’s place, to predict another delivery time and updates it to the customers. The delivery time predict system is able to predict the weather, which means that once it has the weather reports, it is able to upload that into the decision tree that it made earlier and change the possibility of situations about the weather. If the weather report says that there is a high chance of raining on the way that the package across, then all the system need to do is increase the weight of raining. If the training data is following:

| Delivery time | Weather | Main traveling time | Location |
|---------------|---------|---------------------|----------|
| Four days     | Sunny   | Morning             | LA-NY    |
| Four days     | Raining | Morning             | LA-NY    |
| Four days     | Sunny   | Afternoon           | LA-NY    |
| Six days      | Snowing | Morning             | LA-NY    |
| Four days     | Sunny   | Morning             | LA-NY    |
| Five days     | Raining | Afternoon           | LA-NY    |
| Six days      | Snowing | Afternoon           | LA-NY    |

Figure 2. Decision tree machine learning algorithm based on Table 1.
Here another idea as implied: weight. The average number of layers (or levels) from the root to the terminal nodes is referred to as the average depth of the tree. The average number of non-terminal nodes in each level of the tree is referred to as the average breadth of the tree. In general, the average breadth of the tree will reflect the relative weight given to classifier accuracy whereas the average depth of the tree will reflect the weight given to efficiency\[6\]. At this point, the delivery time predict system is increasing the weight of rain, which means that rain becomes more dominant than other factors, which fit the idea that “there will be a high chance of raining”. Then, going through the decision tree, the predicted system will increase or decrease the predicted delivery time using data from a real-time feedback system — both locations and the possibility of the change of weather.

By following that concept as the package is moving towards the location of customers, the predicted delivery time would become more and more precise, which would help customers to have a better idea about when the packages will arrive.

3. **EXAMPLES**

Recently, most of the famous United States companies still have not build a mature system about delivery time. Still, they predict the delivery time based on a simple machine learning algorithm based on all the possible situations that exist to the package on its way to the location of customers. That algorithm is still naive because the predicted delivery time and the real delivery time are different from a number of four to five days. That number is overly big and could not help customers much by knowing that predicted time. But, other companies in other countries have built a mature delivery time predict system. One of the companies that especially need to be accounted for is Taobao, a Chinese company.

Instead of using “IoT and 4PL with soft infrastructure (IT, Human skills and knowledge), hard infrastructure (Tracing and tracking, sophisticated trucks) and flexibility (in both hard and soft infrastructure)”, Taobao is currently using a model that Coltman and Devinney proposed: “a model with operational capabilities for customized and commoditized services. The operational capabilities include customer engagement, cross-functional coordination, creative solutions, operations improvement, IT infrastructure, and professional delivery” \[7\] the professional delivery is one of the parts that other companies are not mature with. Instead of just do one overall predict delivery time, they change the predicted delivery time as the passage is moving from one transfer station to another. They do not use a real-time location feedback system because they would record the location of the package every time it arrives at a transfer station. Doing so is not as accurate as predicting the time as doing it more frequently, but the good thing about the only record the location, once the package arrives at a transfer station, is that they would need fewer data to build their machine learning model. One thing that they do take account of is the use of real-time weather feedback and use it to support the machine learning system and get a better result of delivery time out of it.

In fact, the delivery time predict the system is not unusual. Many other companies other than Taobao, such as Dangdang or Jingdong, have all accomplished the goal of predicting a precise delivering time.

4. **CONCLUSION**

The delivery time predict system will soon be updated on all big companies who consider online shopping and delivering as part of their main businesses. Since this concept is not hard to achieve and this will increase customers’ satisfaction with the companies, this system will soon become the popular predict system across the whole world.

The first problem that this delivery time predicts system is the cost. Even a precise delivery time could increase customers’ satisfaction, the cost on the system and hardware is not a small number. Someone might consider that using a real-time location feedback system on each of the packages is too expensive and ineffective. That is true if each of the packages contain a real-time location sensor. But, since trucks moves the packages, and every time they move hundreds of thousands of packages together, only trucks that contains all packages need real-time location sensors instead of each of the packages.
The result by only put up real-time location feedback system on trucks will reduce the cost while collecting enough data to help the machine learning model to get a better result.

Still, the overall problems that theses kind of system face are data. It is hard to find enough data to train the machine learning model, to begin with. Also, while collecting data by using the real-time location feedback system, how precise the system collects the location data is still affecting the machine learning model. If the system targets the location data from a small area, the data collected would be precise but it needs more data to fulfill the entire model. If the system targets the location data from a big area, it will need fewer data to fulfill machine learning model, but the margin of error in the model would be higher. The problem that needs to be solved is how large should the system target the location data, and that problem is hard to answer because there are positive and negative effects when increasing or decreasing the range, so finding the area in between is hard but important.

Overall, even though there are some shortcomings about the delivery time predict system, this system would be popular around the world soon and give customers a clear idea about when the package will arrive and where the package is now.

ACKNOWLEDGEMENT

Thanks to all my teachers who helped me in this project. This research paper would not exist without their helpful comments.

REFERENCES

[1] T Geraldin, P Giraldin, R Kelly, “Real time location system and method”, US Patent App, 2006
[2] KP. Murphy, “Machine learning: a probabilistic perspective”, Google Books, 2012.
[3] E Alpaydin, “Introduction to machine learning”, Google Books, 2020.
[4] M.A. Friedl, C.E. Brodley, “Decision Tree Classification of Land Cover from Remotely Sensed Data”, Elsevier, 1997.
[5] Edward N. Lorenz, “Empirical Orthogonal Functions and Statistical Weather Prediction”, Bobweigel, 1956.
[6] Safavian, S. Rasoul, Landgrebe, David, “A survey of decision tree classifier methodology”, NASA, 1991.
[7] Joe Yu, Nachiappan Subramanian, Kun Ning, David Edwards, “Product delivery service provider selection and customer satisfaction in the era of internet of things: A Chinese e-retailers’ perspective”, International Journal of Production Economics, 2015.