Considerations

April 11, 2020

Understanding the reasons of delay (and predicting a response?)

Delays at the airport can have two origins: physical and human ones.

Physical reasons

Examples: slow aircrafts, unexpected damages/needs of the aircrafts, closed routes, bad weather conditions..

DATA ANALYSIS

It's the mean by whom we DETECT the reasons of delay.

- **How?** correlation coefficients, distribution analysis, etc. to get what is happening and the causes. TO DO (as soon as possible): we should ask Loic to provide for his own data analysis to be able to focus on aspects that haven't been explored yet. But be careful! Correlation is not causation → the take off movement is correlated to a longer taxi time, but we wouldn't define it as a cause of it.
- **What kind of data do we need?**
  - Schedule data are the most important ones I think because an aircraft that is late tends to wait at the parking slot/at the gate.
  - Data about the length of routes in the airport (this way we can estimate the distance travelled from the aircrafts thanks to the "cheminement" data of our dataset).
  - Weather data.
  - TO COMPLETE
- **What can we do with the data we already have have?** Simple correlation coefficients of temps de roulage with other variables and distributions analysis. This can ureveal relatively simple causes (more complex

causes, derived from the way of handling the ground traffic, can be unrevealed from a more complex model). Loic can provide for some of these results. ← TO COMPLETE

- **Feature Engeneering.** We can generate data from the editing the ones we already have (extrapolate more informations thanks to encoding techniques...). Our work on this part could be really useful, but we need to understand how the airport works to think about features that make sense.
- **To explore:** security distance and maximum speed for different aircraft types.
- **What is the data analysis of reasons of delay useful for?** In our ABM → these situations add a random noise to the take off and landing times that must be included in our ABM to make it resistant to unexpected delays and change of schedule. But to make the model more adherent to reality, we can directly build in our model the structure of the airport and consider as a variable in our model ex. which routes are closed...
- **What did we already discover?** ...

Human and procedural reasons

Deciding that an aircraft has to take of from the northern runway instead of the southern one... → expecially in not standard situations (the ones we're more interested in).

ABM

We can't develop a model to solve the "phyisical" reasons of delay since they are technological problemS. What we can focus on and improve are the procedures followed by GC and pilots to handle the traffic. This is a typical agent based simulation object since it involves decision making and exchange of information between agents. We need to know how does the airport work exactly, i.e. standard procedures that are followed.

The aim of the simulation I have in mind is to see the impact of different policies and decisions taken from the ground controller. In fact, expecially in not standard situations, the eventual delays can be caused from the decision of directing an airplane to nord runway instead of the south one, making him drive along a specific, non-optimal path...

IMPORTANT: we must define a meausure of how good is the behaviour showed by the airport platform agents → can it be just the avarage delay ex. per hour? Are there any other important parameters to account for? (for example in the simulation, the avarage waiting time at the runway doesn't change if we swich from first.in-first-out policy to first-in-last-out policy, but the distribution of delay is really different→in the lifo case we have a really curve of waiting time, that arrives to 20000 seconds of dealy, but less aircrafts have to wait, what is better?)

What can we reproduce with the current ABM? The airport capacity: the airport has a capacity i.e. maximum number of aircrafts that can use the runway in one hour, that depends mainly on the structure of the airport (how many runways does it have, is it well linked to the gates...) and of the airplanes (bigger airplanes move slower and need more security distance). In our model the capacity depends on the time needed to empty the runway-¿ when this time is greater than the avarage time interval between two departures (this time depends on the schedule) we see that long queues take place.

What I think we sould do first

Basically questions for Loic.

- Get the standard procedures of the airport and read them carefully.
- Ask for the stats of Skylab that Loic showed us during the meeting and ask him to explain why he chose those statistics in particular and what are the results (this is one possible starting point to understand the operating modes of the airport).
- Loic spoke about prediction: what does he want us to predict? Temps de Roulage? The general behaviour of the aircrafts in the taxiways?

My ideas

- If it's useful to predict taxi time/delay for every airplane, we could make a machine learning model that has as target variable the taxi time/delay. In this case the feature engeneering part would become really fundamental (we could turn it into a temporal serie problem also).

  PROBLEM: the feature in our dataset can be known only after the landing/take off is finished, while it might be of the airport's interest to forecast the taxi time thanks to informations known before.

  QUESTION: In which phase would they use the taxi time forecast? During the landing/take off? Or while they are deciding the schedule? The data we could use to do the prediction change based on the use they want to do of the prediction.
- Introducing Ground controller strategies in our ABM as explained before.

Further clarifications

The type of data we'll use depend on the usage of our program. Is it for understanding the reasons of delay? Then classical statistical tools may be sufficient to do the job-¿ correlation coefficients... What do the airport staff already know about this?

Is it for forecasting? If it is, what do we want to forecast and in which time horizon? When we will have fixed the output we want to obtain, we can discuss about the inputs we need to obtain it.

For example, if we want to predict the taxi time of an airplane before it starts the taxi process, we have three "classes of availability" for our data. 1) The data "vitesseDeRoulage", "Cheminement", etc. are surely not available. 2) Data about the runway from which the aircraft is going to take off/ the gate it is going to stop at may be available or they may be object of prediction. 3) Data about scheduled take off/ landing and weather data are surely available (in the case of weather data we should have a real-time updating of them).

To summarize: we need to know what the project is for to start analyzing the data and build a program.