DIGITAL ACADEMY WEBINAR

Towards higher levels of automation – Artificial Intelligence (part 2 of 2), 24 July 2020

Questions and answers

Question	Answer
"follow-up my question, so your model has airspace profiles and the data in training model tagged airspaces that were not available in the past (closed) ?"	Exactly, we do (on purpose) a jump in time and use information not available at prediction time, but only during training. On inference phase we use the schedule of course (no other choice :))
@Ramon: do you have any indications about the "systematic errors" you are referring to?	For instance, if the available turn around time is 40 minutes, and the aircraft is expected to depart in 20 minutes, the model could learn that this is impossible and that the take-off time of the ETFMS is not correct (the flight will mostly depart late). Based on the available turn-around time and its complex interaction with other features (e.g, demand, temperature, etc) , the model tries to improve the prediction.
To Mr Belkoura: How did you solve the problems of inter-rater inconsistency (different ATCOs giving different ratings to the same traffic situation) and intra-rater inconsistency (same ATCO giving same rating to vastly different situations)? After model fitting, did you test it on another airspace?	The new experiment will include 20 ATCO per sector for 3 different sectors. This will enable us to make these analysis (sector wise, and ATCO wise). I am expecting different threshold per sector given their changing geometry. On the ATCO consistency, we hope that the aggregated data will show a strong core tendency. We will see soone enough :)
Why did you choose GBDT and RNN as ML algorithms?	GBDT for its simplicity and excellent performance in most state-of-the-art problems using structured data. GBDT is one of the models that wins most of the data science competitions. In addition, it has the advantage that you can "understand" its behaviour.
An additional minor question from mys ide regarding: Can the notion of "systematic error" be further clarified?	yes. By "systematic" I meant prediction errors that can be explained by selecting the right features. For instance, when there is a lot of congestion at a given airport, the ETFMS may be too optimistic and, in most of the cases (systematically), the aircraft departs late.
I have a question for Mr Belkoura on the complexity project: how do you calculate the probabilistic trajectory (the red area)? Thank you very much.	In this phase, we used a ML (more precisely CNN) algorithm to predict gaussian mixture distribution (3 parameters per dimension and 3 mixtures) on the x,y,z axis and a simple cross-entropy on the time axis.
A minor "thought" from my side regarding the collection of big data and the analysis of mobility patterns: Are you also collecting data regarding e.g. planned closure of roads, bridges etc ?	Yes, this kind of natural experiments are indeed one of the major contributions of these opportunistic data sources: learning how behaviour changes in the presence of certain events. For example, we have used this in Madrid to analyse how travel behaviour changes with different meteorological conditions. The next natural step is to use this knowledge to build predictive models that anticipate the impact of future events.

will speech recorgnition will be used to detect alarms, (e.g. a TCAS or ohter warnings coming from display, e.g. by means of searching through ambient voice recording) or a hyterodyne effects on a frequency to detect a simulatious transmission ?	This depends on your application. It is possible. As soon as we have some applications of speech recognition running, the number of applications is very high.
Regarding the flown trajectory project for MUAC, I would like to know more about how the AI models deals with the military reservation areas. Are these areas (as planned by AUP) provided to the system? Thanks	Hi Stella, The AI models actually used the actual opening/closing times of the military areas, during training phase. Because we wanted it to force the algorithm to NOT try to learn uncertainty of military schedule, but consider information as certain. Now, on inference phase, we use the schedule, but the algorithm responds to it as if it were the true one. Is this responding to your question clearly enough?
Wouldn't it be better to replace voice commands of ATCO with text commands, at least for the short once, to keep error rate of interpretation by pilot as low as possible?	yes and no. The short one's are easy to recognize. As soon as we have a speech recognizer in the ops room the number of application are really unlimited. Speech is still the most human natural communication. What about t, that ACTo still speaks to pilot, but the recognition is also sent to the aircraft?
Do consider weather conditions like thunderstorm cells	As of now yes and no :) weather is included in the "hard-coded" logic of the MUAC trajectory pipeline. So in a sense, it is used. However, when we will have access to weather data with more granularity and certainty, we would like to include it.
I am very interested in speech to text recognition. May I have a copy of Hartmut's presentation?	The webinar is being recorded and the recording will be made publically available, together with the presentations.
Would appreciate some comments on how to address two crucial aspects (even if not strictly "technical/scientific") : - Data ownership - Protection of personal data Thanks.	On data ownership: in the case of mobile phone data, the data belong to Mobile Network Operators. We access these data through commercial agreements so that they can monetise their data On data protection: the process comprises the following measures: 1) The raw registers are pseudonymised by the MNO, 2) data are processed in a secured environment at the MNO's facilities, 3) all results are aggregated so that any risk of reidentification is eliminated and the output data are thus fully anonymous