NOTAMATION: ARTIFICIAL INTELLIGENCE AT WORK

The huge number of NOTAMs is a well-known problem in aviation, for pilots at least. What many don't realise is that NOTAMs also present a challenge to EUROCONTROL's Network Management Operations Centre staff. **Camille Uylenbroeck** explains how artificial intelligence will help to reduce the burden.

KEY POINTS

- **Artificial intelligence (AI) is able to help operational staff to manage and prioritise their time and decision-making.**

- **Based on statistical analysis, AI can mimic past decisions.**

- **AI comes with new ways of working that need to be defined with the help of operational staff.**

- **AI does not manage novelty well, so systems must cater for it, by introducing safeguards and by keeping humans in the loop.**

- **AI can be misleading for humans, who must develop new skills.**

This is a story about automation, how we can benefit from it and why we should still be careful about it. It begins with a little notice, called NOTAM (Notice to Airmen), sent to alert aircraft pilots of potential hazards. This notice is very flexible: Many different actors can send it at any time, about anything, and each one of them has its own way of working, which can be very positive.

Unfortunately, it can be very difficult to deal with NOTAMs when you are at the receiving end of so many of them. All NOTAMs are different and are presented in capitals, with obscure coding and

abbreviations. We are not talking here about the aircraft pilots, but about the people working in EUROCONTROL's Network Management Operations Centre (NMOC). They receive between 500 and 900 NOTAMs per weekday and must read them all to find the needles in the haystack – the 1.5% of them that lead to 'data modification requests' (DMRs). These are requests to register NOTAM changes in the EUROCONTROL data. Today, this must be performed continuously, while doing other tasks, never knowing when an important NOTAM is going to show up.

Innovation comes in the picture

A year ago, EUROCONTROL operational staff submitted this problem to an idea collection meeting and my team (the Network Manager Lab) proposed a solution: What if we built a machine learning model able to provide a relevance score to the operational people? What if we built a magnet, to organise this giant haystack and to put the potential needles first?

"What if we built a machine learning model able to provide a relevance score to the operational people?"

After all, we had an enormous amount of data, collected over years, concerning NOTAMs and the subsequent DMRs. This seemed ideal for machine learning. We could teach a classifier to recognise the same features of relevance to NMOC staff to decide whether a NOTAM is of interest. The software would then be able to give a score to each NOTAM accordingly, sorting them to help NMOC to make a decision. And that is what we did.

We talked to NMOC staff, a lot. We tried to understand, as much as possible, what they were watching out for, why they would make one choice or another. We also analysed the data, trying to capture patterns and influences concerning DMRs. We ended up with a mixed approach, using classic algorithms, machine learning

and natural language processing techniques, to understand the human-written NOTAM texts.

Learning from the past

We first allowed the operators to enter keywords to get an artificially high score for the NOTAMs containing the keywords. Artificial intelligence does not manage novelty very well. Events like an eruption of a volcano, flights dedicated to COVID19 vaccines, or closure of an entire country's airspace are very rare. However, we wanted the tool to be as flexible as possible, to be able to gather NOTAMs describing an exceptional situation NMOC wanted to keep visible – one that the machine learning classifier never saw before.

Then, we coded some simple filters to refine the stack before training the machine learning classifier on it. After some data cleaning, we also learned the most frequent sequences of words that *never* produced DMRs, such as 'ACT MUST BE COORDINATED', indicating an airspace activation (usually also notified in other tools, so never needing DMRs). These two steps attribute the lowest possible scores to the NOTAMs, indicating that a DMR is unlikely.
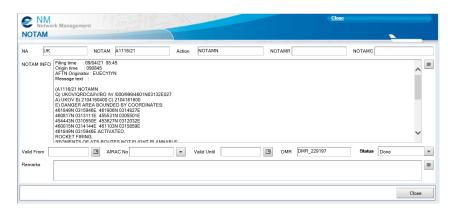
Building the artificial brain

Finally, we used a small 'neural network' as a classifier. A neural network is a set of connected nodes called artificial neurons, which roughly model brain neurons. The network comprised three hidden layers of 10 neurons, to predict a DMR probability for each NOTAM. A little balancing was necessary here, to give more weight to the DMR class, otherwise the algorithm thinks

itself very smart: it just predicts that no NOTAM ever produces a DMR, being right 98.5% of the time. We then transformed this DMR probability into a simple scoring system, from A (the most susceptible to producing a DMR) to F (the lowest possible score).

The most important and the most difficult job was to give to the algorithm the best features to make a choice: the ones playing a role in the DMR probability of each NOTAM. Some features were simple, such as indicating if the NOTAM was a 'new', a 'replace' or a 'cancel'. Some features were obvious but more difficult to engineer, such representing the human-written texts via natural language processing computer programs (designed to process and analyse large amounts of written text).

We ended up choosing the presence or absence of the most discriminating sequences of words. Finally, some features were added thanks only to NMOC staff willingness to work with us and explain – and reexplain – their work. For example, two types of NOTAMs (the 'replaces' and the 'cancels') are applied on previous NOTAMs. It appeared that the presence or absence of DMRs for these NOTAMs were among the most important decision criteria for NMOC. Therefore, this became one of the most important features of the classifier.

"The most important and the most difficult job was to give to the algorithm the best features to make a choice"

The machine sorts while human focuses on the essentials

Finally, after lots of refinements, changes, trials, errors, and corrections… it worked. Just in the 'A' score, we managed to catch 83% of the NOTAMs producing a DMR among the 7 to 14 daily 'A' NOTAMs. If we add the 'B' score, we catch 92% of our needles among the 17 to 31 'A' and 'B' NOTAMs. With the 'C' score, we catch 96% of them among the 45 to 82 A, B and C NOTAMs arriving every day.

To summarise, instead of browsing randomly arrived NOTAMs, an operational staff member will first look for A score or A and B score. With a review of a maximum of 31 NOTAMs, 92% of those requiring a change in NM systems are caught.

We also hope to reduce use errors by allowing operators to take more time for the more relevant NOTAMs. However, the introduction of AI could also produce another source of errors via the so-called automation bias. It happens when users put so much trust into the technical system that they ignore valid contradictory information. For example, a NOTAM labelled E or F will be difficult for the operational staff to link with a DMR, because it will be such a rare event. However, automation bias can be reduced here by decreasing the complexity of the information displayed, like transforming a probability to a letter score, or providing support information instead of directives or commands.

As good as it gets?

This tool will not prevent NMOC from going through the entire daily batch of NOTAMs, because they cannot afford to miss even one NOTAM. Indeed, the software will be wrong sometimes – it is inevitable. However, the tool will help the operational people to have flexibility to prioritise their actions according to the score given to the overdue NOTAMs. Therefore, they will be a lot quicker to treat the NOTAMs with DMRs, since these NOTAMs will be at the top of the list. Staff can then to choose a quiet time of the day to review the rest. This is what we were aiming for. For now, the tool is as good as it gets. **S**

"The tool will help the operational people to have flexibility to prioritise their actions according to the score"

Camille Uylenbroeck has been working in EUROCONTROL as a Sopra Steria data scientist for three years, after graduating in computer science engineering. She has worked among the NM Lab team on several innovative projects involving machine learning. She has also won the Digital Sky Challenge of 2019, a 48h Hackathon, with her team in the environment category.

An operational perspective

"At the Network Manager Operations Centre (NMOC) we optimise traffic flows by constantly balancing capacity and demand, while helping to ensure the safe and efficient operation of flights going to and over Europe. To support this ATFM function, NMOC is monitoring manually between 600 & 800 NOTAMs daily, from which only a small percentage requires an update in our system in real time. This new AI, by assigning priority index on each NOTAM in an interactive dashboard, will reduce significantly the workload in NMOC. It is an important step towards the modernisation of our systems.

"One of the important missions of NM is to ensure the European Crisis Management function. Recent crises (volcanic, pandemic, geo-political) have highlighted how vulnerable our Network can be during these situations. In such cases, NM needs to bring a rapid, well-informed and coordinated response to minimise any negative impact and maintain safe and efficient operations at local and network levels. To support this task and to be a leader of this function in Europe, detailed monitoring of NOTAMs by NMOC is crucial. Integrated in a new HMI NOTAM dashboard, the AI will also allow us to receive in real time, and with high level of priority, any alert about a potential disruption or crisis in the Network."

Daniel Degehet, Operations Manager at EUROCONTROL NMOC

Daniel Degehet was a Belgian Military ATCO for several years before joining EUROCONTROL in 1993 to participate to the creation of CFMU. In 2015, after 20 years' experience in NM Operations, he was promoted to Operations Manager in NMOC.