Machine Learning Model for Aircraft Performances

Marko Hrastovec
CNS
Slovenia Control, Ltd.
Brnik, Slovenia
marko.hrastovec@sloveniacontrol.si

Franc Solina
Faculty of Computer and Information Science
University of Ljubljana
Ljubljana, Slovenia
franc.solina@fri.uni-lj.si

***Abstract*—Aircraft performances are a very important input for trajectory calculations. All known models distinguish performances mainly on aircraft type. We know there are many other factors influencing aircraft performances like operator, aerodrome, take-off weight, etc. We propose to build a machine learning model that would predict aircraft performances based on all known attributes about the flight. The model learns from real flights in the past and makes predictions based on accumulated knowledge. For instance: A full flight to a distant holiday destination has different performances than a short regular flight with the same aircraft. Our model would recognize the difference in that case and predict aircraft performances best suited for each flight.**

aircraft performances; machine learning; data mining

I. Introduction

With the constant increase of air traffic the airspace becomes denser. In order to maintain the level of safety, airspace management faces a great challenge to cope with traffic increase. For that purpose better and more precise planning of flights will be required.

Calculating 4D trajectories for aircraft is becoming more and more important. With the amount of air traffic minimal improvements in flight paths can bring enormous savings in fuel, air pollution, flight delays, airspace optimization, etc. Good knowledge where aircraft is going to be at certain time is very important for avoiding potential conflicts in the air.

Present clearance-based operations rely on air traffic controllers to identify and resolve potential conflicts [1] [2]. This will not be enough any more. We will have to find new ways to manage airspace in a safe way. There are more ways to accomplish the tasks ahead. One is introducing new tools for air traffic controllers, which will enable them to identify potential conflicts earlier. This leads to gradual move towards trajectory based operations and throughput optimizations in positioning aircraft closer to each other. All methods named require exact as possible flight path calculation and prediction.

Very important input for 4D trajectory calculations is Base of Aircraft Data (BADA) model [3] [4]. It is based on physical characteristics of aircraft. Aircraft are grouped in BADA according to type (e.g. Airbus 320, Boeing 747). This classification uses average values for the type to fit general characteristics. Sometimes type is not enough to adequately classify an aircraft, because there can exist various sub-models which have significantly different performances. Every flight is also dependent on many influencing factors like load (take-off weight), weather conditions, geographical environment, etc.

We propose a machine learning model, which takes into account many other factors and predicts aircraft performances according to them.

II. Past Achievements

Base of Aircraft Data (BADA) model is de facto standard for short-term and long-term 4D-trajectory calculations [5]. It is based on mass-varying, kinetic approach that models an aircraft as a point and requires modeling of underlying forces that cause aircraft motion [6]. With the help of complex formulas aircraft performances are calculated from aircraft characteristics provided by manufacturers.

To best of our knowledge there is no BADA alternative publicly available. Currently the new version BADA 4 is being developed which is supposed to provide good model for future needs in aviation flight planning [3].

To improve trajectory prediction accuracy, even more complex models have been developed [7] [8]. They use BADA and data from real flight trajectory recordings. The model by Schuster, Ochieng and Porretta [8] uses flight management system to make flight path more realistic. It combines aircraft performances and flight intent to predict and adjust trajectory accordingly.

BADA also provides a generic aircraft behavior model called AiRline Procedure Model (ARPM), which focuses on how the aircraft is operated [9]. During climbs and descents the energy share factor defines how much of the available power is allocated to vertical evolution as opposed to acceleration.

Gillet, Nuic and Mouillet [9] are using radar recordings to get realistic data to fine tune ARPM. Different energy share factors and speed profiles have been calculated according to airline operator, operating airport, aircraft type, flight phase and flight range with the help of statistical processing. That way they generate more realistic flight trajectories for simulation purposes.

De Leege, van Paassen and Mulder [10] are using machine learning methods to predict trajectories along one particular landing procedure of 45 nautical miles length. Trajectory prediction is predicting time over points from first approach navigation point along significant points to the runway threshold (total of 7 points). Model inputs are: aircraft type (heavy, medium), aircraft ground speed, altitude over initial point and winds. The model predicts with approximately 5s error on the last 15 nautical miles and 20s error on 45 miles trajectory. With model calculated approach schedule, the capacity was increased by four aircraft per hour.

Kun and Wei [11] are similar in the context of ignoring aerodynamics and using radar data. The method consists of two phases. First, they predict total flying time based on historical data of identical flights. The second phase of prediction is adjusting the trajectory based on real-time radar data after the flight takes off.

Cheng, Cui and Cheng [12] use a hybrid of neural networks and statistical analysis. The proposed prediction model was tested on air traffic flow collected by the Air Traffic Control Command Monitoring System (ATCCMS), which aims to give early conflict alert and advice of short-term air traffic flow management to human controllers in the Beijing center. Through the analysis, the air traffic flow was classified into seven categories corresponding to daily difference in a week, which were trained and forecasted separately.

We propose a generic machine learning model which will provide an input for any trajectory calculation method which uses aircraft performances. Calculations should be more accurate with aircraft performances knowledge derived from historical data.

III. Alternative Approach With Machine Learning

We are taking a completely different approach in computing aircraft performances. We do not use an explicit physical model but learn the aircraft performances from data. Learning physical properties and physical laws from data has been tried often [13] [14]. Our approach relies on collecting large amounts of data and enriching them with additional non-physical attributes. For instance, we are not focusing only on flight characteristics. Attributes like company, aerodrome of departure, aerodrome of destination, day or hour of flight, etc. are also influencing the aircraft's performances. We can learn from such attributes that a flight from Ljubljana to Munich is usually full on Monday mornings and it will most probably climb with lower climb rate as the same aircraft on the same route on Tuesday. Another example could be two aircrafts of the same type departing one after another. One is a regular airliner and the other is a charter filled with holiday passenger and their luggage. Using solely an aircraft model we cannot predict that one will climb much faster than the other.

With information about company, destination, etc. we can quickly deduct and predict aircraft performances in a given situation better. Such information cannot be found in any physical model. The big quantity of collected data allows us to group flights on these attributes and extract aircraft performances for them. If the aircraft flies a certain route the same way many times, there is a good chance, it will fly similarly today.

In that way we can use statistical methods to find flights and get their performances without direct knowledge of physical characteristics of aircraft. Good historical data are the basis for good prediction.

A. *Sources of Data*

Air traffic control relies on radars. Radars are positioning the aircraft in the air as accurate as possible and are the eyes of air traffic controllers. Our main source of aircraft performances are radar data recordings. For safety purposes there is never just one radar covering portion of an airspace. So we always get more traces of aircraft in the air. Air traffic control uses software called tracker to combine inputs from all radars into one generated air traffic situation picture. This synthetic picture is combined with information from all radars. In that way measurement errors from single radars are effectively minimized and smooth trajectories are extrapolated.

We use tracker generated data as our radar data source because error correction and smoothing have already been performed.

Another important source of data are flight plans. Every aircraft files in a flight plan prior to departure. With radar’s help only, we would not be able to distinguish aircraft types and other important attributes. The flight plan holds information important for air traffic control like type of aircraft, airline operator, aerodrome of departure, aerodrome of destination and many others. We store all these additional attributes with every aircraft performance recorded. In that way our machine learning model is able to distinguish flights based on many other attributes and not only aircraft type.

The third source of data is weather. We are using information about wind and temperature. We can not know at what actual speed the aircraft is flying without weather information. Trackers are calculating ground speed from radar plots. To estimate air speed we need to take into account wind and temperature. One can imagine that air speed can be 100 knots higher than ground speed if the wind is blowing into the nose of the aircraft and 100 knots lower if wind is blowing from behind. Actual air speed must be calculated from ground speed and estimated weather conditions at the flying altitude.

We are getting the upper atmosphere data from numerical weather predictions (NWP) which are generated by environmental agency. The main source for weather predictions are radiosondes which are deployed every day.

In addition to that we are collecting upper atmosphere data also with the help of Mode-S radars. Aircraft are measuring wind and temperature and approximately 6% of them are reporting them to the ground via Mode-S radar. Since we are getting the data from aircraft all the time it may be more accurate than prediction calculated for 24 hours in advance. Studies are showing that these measurements of good quality [15] [16]. When the measurements from aircraft are available we are using them. We are using values from predictions only when aircraft measurements are unavailable.

B. *Preprocessing*

With all three main data sources described in section III.A we first need to do some preprocessing of the data. This phase is very important because good data for learning leads to better results in the end and vice versa.

First, we must calculate the best possible airspeed estimation. Radar is only capable to measure ground-speed so we must subtract wind speed vector from ground-speed in order to get proper aircraft's airspeed. We are getting winds and temperatures from numeric weather prediction models. In addition to that we can get very good measurements from aircraft from which we can build more precise grid of weather conditions. NWP model prediction comes only for a few predefined points. Actual measurements taken on the aircraft at the exact time of flight seem a better alternative than prediction calculated for many hours in advance. So we are using aircraft measurements whenever possible.

With better airspeed we can proceed to the dissection of the flight. We must find phases of flight where aircraft is ascending, descending or flying on level. Trackers provide that information with the tracks. We record and calculate appropriate performances for different phases. For instance, airspeed can be measured during all flight phases while climb rate only on climbs and descend rate on descents.

When we can extract all aircraft performances from flight, we get simple facts like ascend rate 1000 feet/minute at altitude flight level 200 (20000 feet). Every single flight can provide from one to tens of such facts. It depends what maneuvers the aircraft is performing. If it is a level flight all the time when visible to radar it can be only one airspeed. If the aircraft is departing or landing there are many facts for that flight. These facts alone have no valuable information until we enrich them with the flight plan and weather attributes. This is one of our most important added value features. Good correlation between radar recordings and flight plans is very important because we need to couple the right attributes with corresponding radar data. Air traffic control systems do this important job since we cannot afford to have a radar track wrongly correlated to its flight plan. We have access to this correlated information and we use this data as input for the machine learning.

Using machine learning methods and the computer’s ability to process large amounts of data we can afford to include many other valuable attributes, which influence the aircraft performances.

Other models for aircraft performances are generalizing and grouping aircraft only on aircraft type to simplify models and keep maintainability on a reasonable level of complexity. This is the only feasible way with manually maintained models.

However, we know that aircraft type alone can not be descriptive enough in many cases. Just recently the maintainers of BADA model were asking users for feedback how to solve the issue of same aircraft models with different engines. There is an option to keep them together and use some average values for both aircraft versions or to ask the International Civil Aviation Organization (ICAO) to split the category in two to be able to characterize every group separately.

At this point we can not tell which attributes will be more important or descriptive than another. It is a job of the machine learning model to make that decision, so we are keeping all available data for now. Maybe at later stages we can identify parameters which are not contributing any information and we may drop them later in the process.

We have already preprocessed the data for flights flying through Slovenian airspace in the period between February 2011 and September 2013. We have gathered over 7 million facts about aircraft performances. Preliminary quality checks for facts were already done and they show consistency of gathered facts. These checks also show some pretty obvious differences in aircraft performances of the same aircraft type operated by different operators.

C. Machine Learning

After data is preprocessed, we need to build a machine learning system which is able to accurately predict aircraft performances based on the accumulated data and known attributes about the flight.

The challenge in this phase is to build an effective model on combination of imprecise and precise data with possibly missing attributes. Since the amount of data is big enough, we did not try to eliminate bad data at preprocessing. We expect that these wrong values will not influence the model significantly because there will be much bigger amount of proper data to learn from.

We are using a multidimensional database for machine learning where each attribute about the flight represents a dimension. When searching for the best prediction, we provide the known or significant dimensional values about the flight to the model.

Our data model is composed of over thirty dimensions. When searching for the prediction, the machine learning model will make better prediction if more details about the flight are given.

1) Incremental and Unsupervised Learning

The preprocessing phase is fully automatic. Every night the flights from the previous day are being processed and put into the multidimensional database. In that way the machine learning model has the latest data at its disposal for predictions.

Since the machine learning model uses the newest data every day, it has no sense to investigate the data in advance in order to extract knowledge. When the prediction is required the model searches through the database and finds the best possible matches. This is unsupervised learning.

2) Machine Learning Algorithm

The algorithm for searching the database and making predictions comes from the class of unsupervised methods for association rules. An apriori algorithm makes multiple iterations to get to the best suitable set of similar flights. Traditional application of association rules algorithm is the market basket analysis, which looks for similar purchases in the transaction database to suggest products, a buyer might be interested in [17]. In our case, the a priori algorithm cannot be directly applied. Due to different data we use statistical methods for quantitative association rules [18]. In that way the algorithm finds a representative set of flights which are most similar and will give the best prediction of aircraft performances for a given flight.

IV. CORRECTNES AND EVALUATION

The results of prediction will be tested on real flights. We are calculating the expected trajectory of the flight with BADA performances and with our predicted performances. Both trajectories will be compared with the real flight recordings. We expect to have trajectories closer to reality with our predicted aircraft performances.

V. CONCLUSION

The initial feedback on our goals is positive so that we can conclude that this topic is of interest to many and can help in fuel savings, reduction in flight delays and other important factors. Potential money savings with precise trajectory prediction are definitely worth the effort.

The proposed method is only the first step in the task of getting better trajectory calculations. We expect to get through this novel and completely different approach to the determination of aircraft performances a better insight in a whole range of related problems.

In a way, we are ignoring the explicit physical model and learning based on indirect data about flights. For instance, we don't have the take-off weight but can predict climb well with indirect attributes, which tell whether the aircraft is full or not. In our opinion, the state of the art methods for trajectory calculations that will emerge in the future will use the best from both approaches. The explicit physical model cannot be completely neglected. On the other hand, we are trying to show that other flight attributes in our research are also an important factor and should be taken into account.

REFERENCES

[1] P. Brooker, "4D-Trajectory Air Traffic Management: Are There 'Killer Apps'? – Part 1," *Journal of Navigation,* vol. 65, no. 03, pp. 397-408, 2012.

[2] P. Brooker, "4D-Trajectory Air Traffic Management: Are There 'Killer Apps'? – Part 2," *Journal of Navigation,* vol. 65, no. 04, pp. 571-587, 2012.

[3] A. Nuic, D. Poles and V. Mouillet, "BADA: An advanced aircraft performance model for present and future ATM systems," *International Journal of Adaptive Control and Signal Processing,* vol. 24, no. 10, pp. 850-866, 2010.

[4] D. Poles, A. Nuic and V. Mouillet, "Advanced aircraft performance modeling for ATM: Analysis of BADA model capabilities," in *29th* Digital *Avionics Systems Conference*, 2010.

[5] Eurocontrol, *User Manual for the Base of Aircraft Data (BADA) Revision 3.11,* Eurocontrol, 2013.

[6] A. Nuic, C. Poinsot, M.-G. Iagaru, E. Gallo, F. A. Navarro and C. Querejeta, "Advanced aircraft performance modeling for ATM: enhancements to the bada model," in *24th Digital Avionics System Conference*, 2005.

[7] M.-D. Dupuy and M. Porretta, "Preliminary results for a robust trajectory prediction method using advanced flight data," in *26th Digital Avionics Systems Conference*, 2007.

[8] W. Schuster, W. Ochieng and M. Porretta, "High-performance trajectory prediction for civil aircraft," in *29th Digital Avionics Systems Conference*, 2010.

[9] S. Gillet, A. Nuic and V. Mouillet, "Enhancement in realism of {ATC} simulations by improving aircraft behaviour models," in *29th Digital Avionics Systems Conference*, 2010.

[10] A. de Leege, M. van Paassen and M. Mulder, "A Machine Learning Approach to Trajectory Prediction," in *AIAA Guidance, Navigation, and Control (GNC) Conference*, 2013.

[11] W. Kun and P. Wei, "A 4-D trajectory prediction model based on radar data," in *27th Chinese Control Conference*, 2008.

[12] T. Cheng, D. Cui and P. Cheng, "Data mining for air traffic flow forecasting: a hybrid model of neural network and statistical analysis," in *Intelligent Transportation Systems, 2003. Proceedings. 2003 [IEEE]*, 2003.

[13] I. Bratko, "Qualitative reasoning about control," in *Proceedings of IEEE 2nd International Workshop on Emerging Technologies and Factory Automation (ETFA '93)*, 1993.

[14] J. Žabkar, M. Možina, I. Bratko and J. Demšar, "Learning qualitative models from numerical data," *Artificial Intelligence,* vol. 175, no. 6-10, pp. 1604-1619, 2011.

[15] B. Strajnar, "Validation of Mode-S Meteorological Routine Air Report Aircraft Observations," *Journal of Geophysical Research: Atmospheres,* vol. 117, no. D23, 2012.

[16] M. Hrastovec and F. Solina, "Obtaining Meteorological Data from Aircraft with Mode-S Radars," *Aerospace and Electronic Systems Magazine, IEEE,* vol. 28, no. 12, pp. 12-24, 2013.

[17] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules," in *Proc. 20th Int. Conf. Very Large Data Bases*, 1994.

[18] Y. Aumann and Y. Lindell, "A Statistical Theory for Quantitative Association Rules," *Journal of Intelligent Information Systems,* vol. 20, no. 3, pp. 255-283, 2003.