Machine Learning Based Aircraft Recovery Optimization

Navid Rashedi, Nolan Sankey, Vikrant Vaze
Thayer School of Engineering, Dartmouth College, 14 Engineering Dr, Hanover, New Hampshire, 03755
navid.rashedi.th@dartmouth.edu, nolan.b.sankey.21@dartmouth.edu, vikrant.s.vaze@dartmouth.edu

Extended Abstract

For aircraft recovery problem (ARP), existing exact methods are too time-consuming — constructing and solving the resulting integer linear programming problems requires too much computational time to be operationally useful. Therefore, existing literature focuses on approximation-based approaches such as solution heuristics and problem size reduction methods that provide a trade-off between computation time, problem size and problem complexity. However, most heuristics that are fast enough suffer from poor solution quality.

In this research, supervised machine learning is employed to expedite optimization. The core idea is to try to identify components of optimal solutions to new problem instances by leveraging their similarity with alternative (historical) problem instances presented in the offline model-training phase. Our approach prunes the decision space by using machine learning algorithms that are trained using the network-wide optimal decisions for similar scenarios.

Problem Formulation

First, we employ a commonly used modeling approach for characterizing aircraft recovery decisions, based on the concept of flight strings. A flight string is a sequence of flights operated by a single aircraft of a particular fleet type. Next, we present the sets, parameters and decision variables and the ARP formulation itself.

Sets and Parameters

S_{agt}^{B}: Set of strings corresponding to aircraft group $g \in G$ that begin at airport $a \in A$ before the beginning of period $t \in T$

S_{agt}^{E}: Set of strings corresponding to aircraft group $g \in G$ that end at airport $a \in A$ before the beginning of period $t \in T$

I_{fs}: Indicator parameter (1 if flight $f \in F$ is in string $s \in S_g$, $\forall g \in G$)

N_{agt}^{B}: Number of available aircraft of aircraft group $g \in$ G at airport $a \in A$ at the beginning of period $t \in T$

N_{agt}^{E}: Number of aircraft for aircraft group $g \in$ G that need to be at airport $a \in A$ at the beginning of period $t \in T$

Decision Variables

x_{gs}: The assignment of aircraft group to strings (1 if string s ∈ S is operated by aircraft group g ∈ G)

y_f: Defines the cancelled flights (=1 if flight $f \in F$ is cancelled)

Constraints

$$\sum_{g\in G}\sum_{s\in S_g} I_{fs} \times x_{gs} + y_f = 1, \forall f \in F \qquad (1)$$

$$\sum_{s\in S_{agt}^{B}} x_{gs} \le N_{agt}^{B} \quad \forall g \in G, \forall t \in T, a \in A \qquad (2)$$

$$\sum_{s\in S_{agt}^{E}} x_{gs} \ge N_{agt}^{E} \quad \forall g \in G, \forall t \in T, a \in A \qquad (3)$$

$$x_{gs} \in \{0,1\}, \quad \forall g \in G,\ \forall s \in S_g,\ y_f \in \{0,1\}, \quad \forall f \in F$$

Objective Function

$$Min \sum_{f\in F} y_f \times c_f^{cancel} + \sum_{g\in G}\sum_{s\in S_g} x_{gs} \times [c_{gs}^{operate} + c_{gs}^{delay} + c_{gs}^{spill}]$$

The objective function is the sum of cancellation cost and assignment cost - the latter is the sum of operating cost, spill cost, and delay cost. Constraints 1 ensure that each flight is either covered by a flight string or cancelled. Constraints 2 and 3 ensure flow balance and aircraft availability at the beginning and ending of the recovery horizon respectively

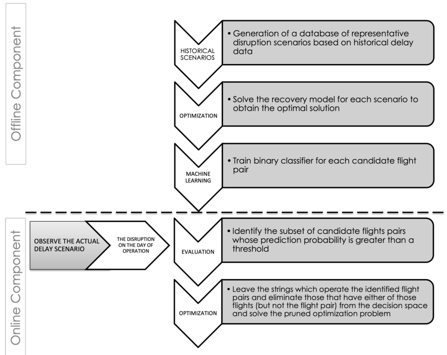

Figure 1 Solution Methodology

Solution Methodology Overview

The methodology consists of two main components: one offline and one online. The offline component should be run before the day of operations. The online component should be run in real-time, between the time when a disruption situation is identified and the time when the airline must decide its recovery actions. The goal is to identify and exploit the similar solution components (flight pairs) between these previously solved offline instances and the online one that manifests on the day of operations, to inform solution of the online instance. A flight pair (FP) is defined as a pair of consecutive flights within a string operated by the same aircraft. In the offline component, first, a set of representative disruption scenarios is identified, and an offline database of disruption scenarios is created. Second, the airline recovery problem is solved for each of these representative scenarios and a solution database is created. Finally, for each FP, a separate classifier, with the aim of predicting the probability that the FP will be selected under a new disruption scenario, is trained. In the online component, first, for the online disruption scenario under consideration, we calculate the probability of each candidate solution component. Then, we identify the subsets of FPs with highest selection probability. Finally, we eliminate from the solution space (of a commercial solver) all the strings which have either, but not both, of those flights. An overview of the solution methodology can be seen in Figure 1.

Computational Results

We present computational case study for JetBlue Airlines during the June-August 2019 time period. Data from BTS (Bureau of Transportation Statistics) Airline On-Time Performance Database were used. We utilized schedules and disruption scenarios from weekdays in June, July and first two weeks of August for training and schedules and disruption scenarios from the weekdays in the last two weeks of August for testing. We consider a day-long recovery horizon, which is typical for many airlines. Flight network consists of 1211 flights with 220 aircrafts (2 aircraft models: Airbus A220 and Embraer 190) and 11 maintenance stations. Features for classification were binary variables defining whether or not a flight was included in the day's flight schedule (acknowledging the fact that flight schedules often change from day to day even during the same scheduling season) and the associated Independent Arrival delays (Lan et al. 2006) of the included flights. Four relevant benchmarks were employed. GS is the *Global Solution* obtained by using the entire string set. TS is the *Trivial Solution* that takes only the disrupted aircraft into consideration. This solution resembles the immediate solution that an airline controller would find during operations as described in Vink et al. (2021). SA is the *Selection Algorithm* – a heuristic that iteratively considers a selection of aircraft, as described in Vink et al. (2021). CGR is the *Column Generation Heuristic* at root node. CGA is the *Column Generation algorithm* at all nodes. ML_LR is our proposed machine learning method using logistic regression classifier for learning and ML_RF is our machine learning method using random forest classifier. Table 1 shows the optimality gap – with respect to the true optimal obtained by solving the problem directly using CPLEX over a much longer time period – of all the approaches for different time thresholds. We notice that at all time thresholds, one of our machine learning-based approaches achieves the lowest optimality gap. Specifically, we establish the superiority of our machine learning-based approaches compared to the other baselines considered for each time budget whether we need to make a very quick decision (30 sec objective) or when the decision can be taken with a maximum of two minutes of available runtime budget (120 sec objective).

Approach	30 sec Runtime	60 sec Runtime	120 sec Runtime
GS	10.6	10.6	10.6
TS	7.2	6.3	6.3
SA	7.2	5.5	4.7
ML_LR	6.5	4.2	3.4
ML_RF	8.3	5.2	2.5
CGA	16.6	12.3	9.1
CGR	9.4	7.5	5.9

Table 1 Comparing optimality gaps of approaches w.r.t true optimal cost at different time thresholds (averaged over 20 test instances)

Conclusion

Our results, demonstrate the consistently high performance of our approach on the tested instances when compared to state-of-the-art approaches under various runtime budgets. While much more exploration, experimentation and close scrutiny is warranted of these machines learning based approaches, we find that they do exhibit a potential to generate fast and high-quality solutions of complex recovery optimization problems faced by airlines.

References

Vink, J. et al. 2020. Dynamic aircraft recovery problem - An operational decision support framework Computers & Operations Research, 117, p.104892.

Lan S, Clarke JP, Barnhart C. Planning for robust airline operations: Optimizing aircraft routings and flight departure times to minimize passenger disruptions. Transportation science. 2006 Feb;40(1):15-28.