

WestminsterResearch

http://www.westminster.ac.uk/westminsterresearch

Airline disruption management with aircraft swapping and reinforcement learning

Hondet, G., Delgado, L. and Gurtner, G.

A a paper presented at the 8th SESAR Innovation Days, Salzburg 03 - 06 Dec, 2018, SESAR.

It is available from the conference organiser at:

https://www.sesarju.eu/sites/default/files/documents/sid/2018/papers...

The WestminsterResearch online digital archive at the University of Westminster aims to make the research output of the University available to a wider audience. Copyright and Moral Rights remain with the authors and/or copyright owners.

Whilst further distribution of specific materials from within this archive is forbidden, you may freely distribute the URL of WestminsterResearch: ((<u>http://westminsterresearch.wmin.ac.uk/</u>).

In case of abuse or copyright appearing without permission e-mail repository@westminster.ac.uk

Airline disruption management with aircraft swapping and reinforcement learning

G. Hondet, L. Delgado, G. Gurtner

École nationale de l'aviation civile

December 5, 2018

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 - のへで

Introduction

Introduction

Disruption management with

reinforcement learning

- Simulator
- Q learning a solver
- Experiments and results
- Conclusion

- Lower costs due to airline disruptions
- Usually, Disruption solution man made by rule of thumb
- Aircraft or flight swapping
- Reinforcement learning

Current work

Introduction

Disruption management

with reinforcement learning

Simulator

Q learning a solver

Experiment: and results

Conclusion

- J. Clausen *et al.* "Disruption management in the airline industry concepts, models and methods", 2009
- R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2017
- V. Mnih et al., "Playing Atari with deep reinforcement learning", 2013

Work done here

Machine learning technique to discover interesting swap combinations

Outline

Disruption management with reinforcement learning

> Hondet, Delgado, Gurtner

Introduction

Simulator

Q learning a solver

Experiments and results

Simulator

Specification Mechanisms Cost & calibration

2 Q learning as solver

Principle and method

Q learning algorithm and implementation Practical training with the simulator

3 Experiments and results

Experimental setup Results

Outline

Disruption management with reinforcement learning

Hondet, Delgado, Gurtner

Introduction

Simulator

Specification Mechanisms Cost & calibratio

Q learning a solver

Experiments and results

Conclusion

1 Simulator Specification

Mechanisms Cost & calibration

Q learning as solver

Principle and method

Q learning algorithm and implementation Practical training with the simulator

3 Experiments and results

Experimental setup Results

> Hondet, Delgado, Gurtner

Introduction

Simulator Specification Mechanisms Cost & calibration

Q learning as solver

Experiments and results

Conclusion

Purpose

- Evaluate the delay on a fleet, on a day of operation
- estimate generated costs
- perform actions on the fleet

Does

- model reactionary delay
- include other delays as probability distributions
- simulate aircraft swapping and its consequences

Does not

- model crew management, nor passengers flow
- manage stand-by aircraft,

Specification of simulator

modify or cancel legs

> Hondet, Delgado, Gurtner

Introduction

Simulator Specification Mechanisms Cost & calibrati

Q learning solver

Experiment and results

Conclusion

Mechanisms

Timestep

 $\forall i \in [\![1, m]\!], t_i$ is the time of the i^{th} landing of the day

 (t_1, t_2, \ldots, t_m)

Actions

Allow to alter the simulation,

"swap with aircraft a"

Cost Immediate cost of a swap

"swapping with a costs c"

Cost & calibration

Cost of what

Delay at departure of the flight after swap

Characteristics

- non linear
- increasing derivative

$$c(d_1 + d_2) > c(d_1) + c(d_2)$$

• depends on the aircraft type

Calibration

Calibrated against Eurocontrol "Coda Digest 2017"

Disruption management

with reinforcement learning

Cost & calibration

Outline

Disruption management with reinforcement learning

> Hondet, Delgado, Gurtner

Introduction

Simulator

Q learning as solver

Principle and methor Q learning Practical training

Experiments and results

Conclusion

Simulato

Specification Mechanisms Cost & calibratio

2 Q learning as solver

Principle and method Q learning algorithm and implementation Practical training with the simulator

3 Experiments and results

Experimental setup Results

Principle

Disruption management with reinforcement learning

> Hondet, Delgado, Gurtner

- Introduction
- Simulator
- Q learning as solver Principle and method
- Q learning Practical training
- Experiments and results
- Conclusion

Reinforcement learning

- Interaction between an agent and its environment
- Find a policy π : state \rightarrow action



Figure: Reinforcement learning principle

Theoretical basis

State $s \in S$, action $a \in A$.

Maximised value

Introduction

Disruption management

with reinforcement learning

Simulator

Q learning as solver Principle and method

Q learning Practical trainin

Experiments and results

Conclusion

$\mathbb{E}\left(\sum_{t=0}^{T_f} r_t\right) \longleftrightarrow Q(s, a) \tag{1}$

Bellman equation

$$Q^{*}(s, a) = r(s, a) + \sum_{s' \in S} p(s'|s, a) \max_{a'} Q^{*}(s', a')$$
(2)

- Dynamic programming
- Monte Carlo simulations

Q learning algorithm

Introduction P Simulator Q learning as solver Principle and method **Q learning** Experiments and results Conclusion

Disruption management

with reinforcement learning

procedure Q-LEARNING(Q) $s \leftarrow \text{initial state}$

while episode not finished do

```
a \leftarrow choose an action from a set
```

```
play a, observe reward r and new state s'
```

```
Q \leftarrow \text{update } Q \text{ with } (s, a, r, s')
```

s ← s'

end while end procedure

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 のへぐ

> Hondet, Delgado, Gurtner

Introduction

Simulator

Q learning as solver Principle and meth Q learning Practical training

Experiment and results

Conclusion

Lookup table implementation

$$Q = \begin{bmatrix} Q(s_0, a_0) & Q(s_0, a_1) & \cdots \\ Q(s_1, a_0) & Q(s_1, a_1) & \cdots \\ \vdots & \vdots \\ & Q(s, a) \end{bmatrix}$$

Update formula

State s, action a, reward r and next state s'.

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left(r + \max_{a'} Q(s',a') - Q(s,a) \right)$$
(3)

Choosing an action

Bandit methods

Maximise reward, minimise regret

Upper confidence bound

 $Q_t(s, a) +$ exploitation

Disruption management with reinforcement learning

Q learning

Choosing an action

Disruption management with reinforcement learning

> Hondet, Delgado, Gurtner

Introduction Simulator Q learning as solver Principle and method Q learning Practical training

Experiment and results

Conclusion

Bandit methods

Maximise reward, minimise regret

Upper confidence bound

$$\underbrace{Q_t(s,a)}_{\text{exploitation}} + c \underbrace{\sqrt{\frac{\ln t}{N_t(s,a)}}}_{\text{exploration}}$$

・ロト・日本・日本・日本・日本・日本

(4)

Final algorithm

Introduction

Disruption management

with reinforcement learning

Q learning as solver Principle and methe Q learning Practical training

Experiment and results

Conclusion

procedure Q-LEARNING(Q, c, α, A) $s \leftarrow \text{initial state}$

while episode not finished do

$$\begin{array}{l} a \leftarrow \text{CHOOSEACTION}(\mathcal{A}, c) \\ (r, s') \leftarrow \text{SIMULATIONSTEP}(s, a) \\ Q(s, a) \leftarrow Q(s, a) + \alpha \big[r_t + \max_{a' \in \mathcal{A}} Q(s', a') - Q(s, a) \big] \\ s \leftarrow s' \end{array}$$

end while end procedure

◆□→ ◆□→ ◆三→ ◆三→ ○○ ◇◇◇

> Hondet, Delgado, Gurtner

Introduction

Simulator

Q learning as solver Principle and metho Q learning Practical training

Experiment and results

Conclusion

Hyperparameters

- Exploitation exploration trade off
- Initial *Q* value

Learning rate

$$\sum_{n\geq 0} \alpha_n = \infty; \quad \sum_{n\geq 0} \alpha_n^2 \in \mathbb{R} \quad (5) \qquad \qquad \alpha_n = \frac{1}{N_t(s, a)} \tag{6}$$

Implementing the training

Chaining training sessions

 $Q^1 \xrightarrow{\text{training}} Q^2 \xrightarrow{\text{training}} \cdots \xrightarrow{\text{training}} Q^*$

▲□▶ ▲圖▶ ▲圖▶ ▲圖▶ ▲□▶

(7)

-1

State space

Observation

Partial information of the environment, $\mathcal O$ the set of observations,

$$(\mathcal{S}, \mathcal{A}) \xrightarrow{\phi} (\mathcal{O}, \mathcal{A}) \xrightarrow{Q} \mathbb{R}$$

Choice of ϕ

- Carries enough information
- But not too specific
- Time independent

management with reinforcement learning

Disruption

Delgado Gurtnei

Q learning as solver Principle and meth Q learning Practical training

> Experiment: and results

> Conclusion

Outline

Disruption management with reinforcement learning

> Hondet, Delgado, Gurtner

Experiments and results

U Simulat

Specification Mechanisms Cost & calibration

Q learning as solver

Principle and method

Q learning algorithm and implementation Practical training with the simulator

3 Experiments and results Experimental setup Results

> Hondet, Delgado, Gurtner

Introduction

Simulator

Q learning a solver

Experiments and results Experimental setup Results

Conclusion

Experimental setup

- Schedule: Vueling, October 12, 2014
- 6 aircraft, 14 stations, 35 flights

Observation

Two different observations tested.

Disruption

Artificial delay added.

Hyperparameters

 $(p_d, c, q_i) = (0.06, 10, -90000)$

Output format

Disruption management with reinforcement learning

> Hondet, Delgado, Gurtner

- Introduction
- Simulator
- Q learning a solver

Experiments and results Experimental setup

Conclusion

Spreadsheet like parquet files

Columns

- delays
 - atfm delay
 - departure delay
 - miscellaneous delays
 - reactionary delay
 - artificial delay added
 - taxi time
- action and reward
 - action number
 - swap or not
 - cost
 - cumulative reward

- simulation information
 - departure destination
 - departure origin
 - leg duration
 - departure sobt
 - tail number
 - tail number of swapped aircraft
 - time in the simulation
- Q learning data
 - state-action couple visit count
 - Q value

Learning process

▲□▶ ▲□▶ ▲ □▶ ▲ □▶ ▲ □ ● ● ● ●

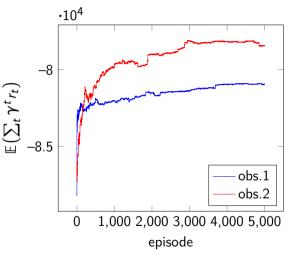


Figure: Average maximum Q values over 5000 episodes.

Disruption management with reinforcement learning

> Hondet, Delgado, Gurtner

Introduction

Simulato

Q learning solver

Experiments and results Experimental set Results

Conclusion

> Hondet, Delgado, Gurtner

Introduction

Simulato

Q learning solver

Experiments and results Experimental set Results

Conclusion

Comparing with idle behaviour

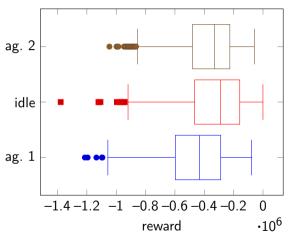


Figure: Comparing the idle behaviour with the agent.

Conclusion

Re

Simulator

Disruption management

with reinforcement learning

Q learning a solver

Experiments and results

Conclusion

Results

Cost reduced in some conditions, not reliable enough. Potential lines of research.

Perspectives

- refine observations
- more sophisticated reinforcement learning techniques
- develop further the simulator