Transportation Research Part E 95 (2016) 282–298

Contents lists available at ScienceDirect

Transportation Research Part E

journal homepage: www.elsevier.com/locate/tre

An analysis of Brazilian flight delays based on frequent patterns

Alice Sternberg [a], Diego Carvalho [a], Leonardo Murta [c], Jorge Soares [a,b], Eduardo Ogasawara [a,*]

[a] Federal Center of Technological Education of Rio de Janeiro - CEFET/RJ, Brazil
[b] State University of Rio de Janeiro - UERJ, Brazil
[c] Fluminense Federal University - UFF, Brazil

ARTICLE INFO

Article history:
Received 23 March 2016
Received in revised form 31 August 2016
Accepted 30 September 2016

Keywords:
Flight delays
Association rules
Commercial aviation
Delay reasons

ABSTRACT

In this paper we applied data indexing techniques combined with association rules to unveil hidden patterns of flight delays. Considering Brazilian flight data and guided by six research questions related to causes, moments, differences, and relationships between airports and airlines, we evaluated and quantified all attributes that may lead to delays, showing not only the main patterns, but also their chances of occurrence in the entire network, in each airport and airline. We observed that Brazilian flight system has difficulties to recover from previous delays and when operating under adverse meteorological conditions, delays occurrences may increase up to 216%.

1. Introduction

Delays are one of the greatest challenges to transportation systems. Notably, in commercial aviation, delay is usually defined by the difference between scheduled and real times of departure or arrival (Wieland, 1997). Despite some differences in tolerance thresholds for delays, country regulatory authorities usually monitor delays levels through several indicators. In 2014, 16.6% of flights delayed by more than 15 min in Europe and 24.7% in the United States. In Brazil, 19.1% of domestic flights were canceled or suffered delays greater than 30 min (EUROCONTROL, 2015; The Unites States Department of Transportation, 2015; ANAC, 2015b).

Flight delays impact passengers, airlines, and airports, especially increasing trip and operations costs. Given the uncertainty of their occurrence, passengers usually plan to travel earlier to ensure their arrival on time. On the other hand, airlines may have to pay penalties, fines, or incur extra cost, such as crew reschedules and aircraft retentions in airports (Britto et al., 2012). Moreover, delays are also related to environmental damages, since they may increase fuel consumption and gas emissions (Pejovic et al., 2009; Ryerson et al., 2014; Simić and Babić, 2015).

Delays also affect the airlines marketing strategies, since the loyalty of customers are motivated by punctual performances (Vlachos and Lin, 2014). Furthermore, delay levels are not only related to operational and economic choices of an airline (such as aircraft sizes, flight frequencies, and fares), but also with complaints about airline service (Bhadra, 2009; Pai, 2010; Zou and Hansen, 2014). In this context, understanding the reasons for flight delays occurrences can direct public and private investments in air transportation systems, improve tactical and operational decisions of airports and airlines managers, and warn passengers, so they can rearrange their plans (Marsden, 2002; Lv and Wang, 2009).

Every moment, a massive amount of data from commercial aviation is collected and stored in public and private databases. Seeking to understand the air transportation ecosystem, domain analysts and data scientists are intensifying the usage

* Corresponding author.
E-mail address: eogasawara@ieee.org (E. Ogasawara).

http://dx.doi.org/10.1016/j.tre.2016.09.013
1366-5545/

of computational and data management techniques to produce useful knowledge from this huge volume of data. Under this scenario, several papers used different methods to predict or understand reasons and behavior of flight delays in the last years. In this paper, we adopt an original approach in the context of flight systems by using frequent pattern mining (association rules learning, more precisely) (Han et al., 2011). We seek to understand the conditions that lead to flight delays discussing several dimensions, such as spatial, temporal, meteorological, features, and previous state of the system. We target to answer the following research questions: (i) What are the main causes of flight delays? (ii) When delays have more chances to occur? (iii) Is there any difference among main airlines in terms of flight delays? (iv) Is there any difference among main airports in terms of flight delays? (v) Is there any relationship between airports and airlines in terms of delays? (vi) When a departure delay does not lead to an arrival delay? The analysis is conducted by a mining process composed of data indexing, rules generation, and rules analysis.

Moreover, most of previous studies concerning flight delay arena focus on evaluating air transportation systems of the United States, Europe, and China (Reynolds-Feighan and Button, 1999; Mueller and Chatterji, 2002; Tu et al., 2008; Lu et al., 2009). Differently than the mainstream, we conducted such innovative analysis on Brazilian domestic flight system to better comprehend its flight delays, their causes, and hidden patterns. Brazil is an important market for commercial aviation due to its geographical and infrastructure characteristics. In 2014, it was the first in Latin American country on the Revenue Passenger-Kilometer (RPK) ranking, being the fifteenth place worldwide (ICAO, 2015). In spite of the importance of commercial aviation in Brazil and the existence of public databases, the knowledge about flight delays in Brazil is still empirical and lacks a broader analysis.

We observed that the original usage of association rules learning to understand flight delays is an interesting approach since our previous statistical analysis of flight delay data did not show, except for departure and arrival delays, any strong correlations between main attributes. Thus, classical statistical analysis, such as regressions, would hardly find out hidden patterns that could answer our research questions. On the other hand, in addition to show such patterns, association rules also quantify the conditional probabilities of the occurrences of consequences (in our case delays) given a set of causes. This is done by an iterative constructive pattern search algorithm that is guided by concepts such as support and confidence (Hipp et al., 2000). Among the main findings in our study, we highlight the influence of fog, thunderstorms, rain, and instrument flight rules in flight delays. We have observed that Brazilian airports and airlines have difficulties on recovering from previous delays. We have also observed that meteorological and temporal attributes are not equally related to delays in different Brazilian airports.

Besides this introduction, the remainder of this paper is structured as follows. Section 2 presents a background of flight delay analysis and data mining. Section 3 presents an overview of Brazilian commercial aviation and delays. Section 4 describes the data mining process used to conduct this analysis, including data indexing techniques and association rules generation. In Section 5 we answer our research questions, evaluating Brazilian flight delays through association rules, showing the most delayed airports and airlines, and discussing the conditions that lead to these delays. Finally, Section 6 concludes our analysis presenting major highlights and insights on how to attenuate the Brazilian flight delay problem.

2. Background

The general background encompasses both flight delay analysis and data mining practices. Due to this, it is divided accordingly in two subsections. Subsection 2.1 introduces the flight delay scenario, describing a typical operation of a commercial flight, kinds of delays, and their impacts. Subsection 2.2 discusses associations rules and the common measures used to filter and evaluate interesting rules, giving a didactic example of their concepts and usage in the air transportation system.

2.1. Flight delay analysis

Commercial aviation systems are often very complex. They deal with expensive resources, demand fluctuations, and an intricate network of airports that may need different conditions to provide smooth and safe operations. Moreover, passengers follow various itineraries, which may differ from the schedules for aircraft, pilots, and flight attendants (Ball et al., 2007).

Basically, a typical operation of a commercial flight starts at terminal areas and runways of the departure airport, passes through airspace, and ends at runways and terminal areas of the arrival airports, being susceptible to different kind of delays. Some examples include mechanical problems, ground handling, weather conditions, air traffic control, runway queues, and capacity constraints (Reynolds-Feighan and Button, 1999; Hunter et al., 2007; AhmadBeygi et al., 2008).

Operations are repeated several times throughout the day for each flight in the system. Due to legal rests, duties, and maintenance plans, aircraft, pilots, and flight attendants may follow different itineraries. So, any disruption may delay the subsequent flights of an airline (Abdelghany et al., 2004). Furthermore, disruptions may generate congestion in airspace or in other airports, creating queues and delaying some flights from other airlines (Xu et al., 2005; Pyrgiotis et al., 2013).

Due to their importance, flight delays have been a frequent subject among papers about commercial aviation. Analyses consider different dimensions and attributes and apply different methods to understand or predict flight delays suffered by airports or airlines. For example, weather impacts on flight delays were investigated through a simulation model considering airport capacities under instrument and visual meteorological conditions (Schaefer and Millner, 2001). Considering different durations of good or bad weather, authors could understand departure and arrival queuing delays. They concluded