

Master thesis/research project topic

Understanding flight delays

Context

Flight delays are a common hassle for many travelers every day. Just in 2017 until now around 20% of the flights in the US have been delayed. Delays are not only inconvenient to the passengers but also increase the airplanes' fuel consumption. Machine learning algorithms have been used in different forms to produce predictions of flight delays, but these can vary for each region of the globe and do not provide thorough insight into the mechanism of the delays.

Association rule mining is concerned with obtaining rules that show an association between input data and a target feature. Just mining all the possible association rules gives rise to a pattern explosion, i.e., there are more rules than we can inspect. To this end, alternative measures of interestingness and subgroup discovery have been proposed. The first focuses on mining rules that fit with what we define to be an interesting rule, while the second focuses on mining rules that differ from the overall distribution of the data and looking at local effects.

Goal

The goal of this work is to apply existing association rule mining and/or subgroup discovery techniques to a regional flight dataset. A region of interest should be chosen, e.g., India, and then the gathered data should be explored in order to better understand the mechanisms that cause and affect flight delays, such as airlines, routes, airports, weather, etc. This should result in useful insights like the chance of delay given certain characteristics.

The initial task is mostly exploratory, but more questions can be investigated after exploring the data. Most flight data is available online and we also have access to other, non-free, sources of flight data.

Research questions

Example research questions include:

- What are the causes of flight delays?
- What are the chances of flight delays given certain patterns?
- Which main airports or airlines have more chances of flight delays?
- How do previous delays affect future delays in an airport?
- How does a certain airport/airline vary from the overall distribution of all airports/airlines?

Realisation

- Literature study
- Select a region of interest.
- Select suitable methods for analysis of flight delays (Association rule mining, subgroup discovery, interestingness measures).
- Execute several methods, and select the most adequate.
- Evaluate the results, derive recommendations.
- Write report / thesis.

Student profile

Good understanding of data mining problems and algorithms; experience/interest in pattern mining and data structures; preferable programming experience in Python or R.

Supervisors

Hugo Manuel Proença (h.manuel.proenca@liacs.leidenuniv.nl)

Matthijs van Leeuwen (m.van.leeuwen@liacs.leidenuniv.nl)

Relevant literature

Overview of frequent pattern mining

[1] Charu Aggarwal and Jiawei Han. *Frequent Pattern Mining*, Springer, September 2014.

<http://charuaggarwal.net/freqbook.pdf>

Frequent pattern mining for Flight delays:

[2] Sternberg, Alice, et al. "An analysis of Brazilian flight delays based on frequent patterns." *Transportation Research Part E: Logistics and Transportation Review* 95 (2016): 282-298.

<http://www.sciencedirect.com/science/article/pii/S1366554516301740>

Overview of Subgroup discovery:

[3] Atzmueller, Martin. "Subgroup discovery." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 5.1 (2015): 35-49.

<http://onlinelibrary.wiley.com/doi/10.1002/widm.1144/full>

Mining Diverse sets of rules:

[4] van Leeuwen, M. & Knobbe, A.J. "Diverse Subgroup Set Discovery." In: *Data Mining and Knowledge Discovery*, special issue ECML PKDD'11, vol.25(2), pp 208-242, Springer, 2012.

http://patternsthatmatter.org/pubs/2012/diverse_subgroup_set_discovery-vanleeuwen,knobbe.pdf

Recent application of subgroup discovery to science of materials:

[5] Goldsmith, Bryan R., et al. "Uncovering structure-property relationships of materials by subgroup discovery." *New Journal of Physics* 19.1 (2017): 013031.

<http://iopscience.iop.org/article/10.1088/1367-2630/aa57c2/meta>