

Modélisation du risque en transport aérien

soumis par l'IATA

Contexte

Le système GADM (« Global Aviation Data Management ») de l'IATA fournit des indications pour une gestion proactive de la sécurité aérienne. En particulier, la base de données IDX (« Incident Data Exchange ») contient des rapports sur des incidents classés selon la taxonomie de l'IATA. L'IDX fournit aux compagnies aériennes un outil d'analyse comparative permettant à une compagnie d'évaluer sa performance de sécurité à l'échelle globale et régionale : la compagnie peut même évaluer sa performance pour une période donnée, un type d'aéronef donné, etc.

Objectifs

L'IATA désire trouver des identificateurs de risque basés sur les données et collaborer avec des chercheurs dans ce domaine et dans celui des applications de l'apprentissage automatique (voir la première référence ci-dessous).

Défi

Il s'agit de définir un modèle de risque qui puisse détecter des précurseurs d'incidents ou anomalies (par rapport aux tendances précédentes).

(a) Nous voulons affecter une valeur de risque à un attribut (un type d'avion ou un aéroport, par exemple) ou une combinaison d'attributs. Ceci nous permettra de détecter des anomalies et de demander à un analyste d'évaluer la situation.

(b) Nous voulons aussi trouver une manière de prédire combien d'incidents se produiront pendant un mois donné, pour un type précis d'incident. Nous avons besoin d'un intervalle de confiance, pas seulement d'une estimation. En d'autres termes, si le jeu de données contient un nombre d'incidents pour le mois de février qui n'appartient pas à cet intervalle, la situation devrait être évaluée par un analyste.

État de la recherche

Jusqu'ici nous avons concentré notre attention sur les modèles linéaires généralisés (MLG), parce que le rapport de taux d'incidents et le rapport de cotes peuvent être déduits de ces modèles et utilisés pour la modélisation du risque. La régression de Poisson ne convient pas parce qu'elle n'a pas passé le test de Cameron-Trivedi. Nous avons donc décidé d'utiliser un modèle de Poisson généralisé (voir la deuxième référence ci-dessous) mais ce modèle avait des problèmes de convergence. Nous avons aussi examiné le modèle binomial négatif.

Étant donné ces recherches, nous avons deux buts précis.

(a) trouver des métriques de risque pour des attributs de nos données qui ne soient pas des rapports de cotes

(b) trouver de meilleurs modèles de prévision de séries temporelles (les MLG peuvent prévoir des séries temporelles mais ne sont pas aussi bons que des modèles de séries temporelles saisonnières ou des modèles d'apprentissage profond)

[1] F. Bati and L. Withington. 2019. "Application of Machine Learning for Aviation Safety Risk Metric." Paper presented at the *17th IEEE International Conference on Dependable, Automatic and Secure Computing (DASC 2019), Fukuoka, 2019*

[2] F. Famoye, J. T. Wulu and K. P. Singh. 2004. "On the Generalized Poisson regression model with an application to accident data." *Journal of Data Science* 2: 287-295.