

Estudo de Padrões em Acidentes de Trânsito no Reino Unido

Pattern study in car accidents in the United Kingdom

RESUMO

João Marcelo Tozato
tozato@alunos.utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil

Fernando José Antônio
fjantonio@utfpr.edu.br
Universidade Tecnológica Federal do Paraná, Cornélio Procópio, Paraná, Brasil.

Uma grande quantidade de mortes ocorre em todo o mundo em decorrência de acidentes de trânsito, nesse sentido, neste trabalho temos como objetivo levantar relações e comportamentos relacionados a este tipo de acidente. Para isso, foram utilizados dados públicos coletados pelo governo do Reino Unido, realizando análises estatísticas combinadas com técnicas de programação utilizando a linguagem *Python*. Evidenciou-se correlações e tendências ao plotar e analisar as mais diversas combinações dos dados, tais como no número de acidentes e acidentados com o passar dos anos, número de veículos envolvidos em acidentes, além da tentativa de desenvolver um modelo que auxilie na predição da gravidade de determinado acidente com base em suas principais características. Pôde-se confirmar as relações examinadas ao comparar as medidas governamentais tomadas mencionadas em relatórios oficiais no período estudado com os resultados obtidos. Além do mais, observou-se uma transição entre dois regimes alométricos entre a quantidade de veículos e acidentados e também uma performance aceitável no classificador de severidade proposto.

PALAVRAS-CHAVE: Física estatística. Acidente de trânsito. Aprendizado de Máquina.

Recebido: 19 ago. 2019.

Aprovado: 01 out. 2019.

Direito autorial: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



ABSTRACT

There is a great number of deaths all around the world because of car accidents, thus this present study aims on raising relations and behaviors related to such kind of casualty. In order to do it, public data collected by the United Kingdom government, statistical analysis and Python programming techniques were used. Plotting and analyzing the different combinations of the data brought up correlations and tendencies, such as the number of accidents and casualties over the years, the total of vehicles involved in accidents, likewise the attempt of developing a model that assists in the severity's prediction of a certain accident based on its main characteristics. By comparing the results with governmental policies taken during the studied period stated in official reports we were able to confirm some of the relationships found. The number of vehicles and the number of casualties revealed a transition between two allometric regimens and a reasonable performance in the classifier proposed.

KEYWORDS: Statistical physics. Traffic accidents. Machine learning.

INTRODUÇÃO

Segundo a Organização Mundial da Saúde (2015, p. 2), acidentes de trânsito representam uma grande parcela de mortes ao redor do mundo: cerca de 1,25 milhão de mortes por ano, além de ser a principal causa de morte de indivíduos entre 15 e 29 anos. Apesar do aumento da população, da quantidade de veículos, da quantidade de pessoas habilitadas e dos preços dos combustíveis na última década, a quantidade de fatalidades no trânsito se estabilizou (*DEPARTMENT FOR TRANSPORT*, 2015, p. 3), porém, ainda está entre as dez principais causas de morte no mundo (*WORLD HEALTH ORGANIZATION*, 2018).

De acordo com o relatório do Departamento de Transporte do Reino Unido (2015, p. 2), foram 186 mil vítimas envolvidas em acidentes neste ano, uma redução considerável de 42 por cento, ao considerar os 321 mil acidentados trinta anos antes. Este decaimento deve-se a uma série de fatores, tais como: melhorias nas medidas de segurança no trânsito, nas estradas e nos veículos do país.

Conforme Clarke et al. (1998, p. 1), a investigação de dados de acidentes e os seus fatores apresentam diversas complicações, tendo em vista a natureza destes que são relativamente imprevisíveis e não constantes, dessa forma, mostra-se bastante necessária no sentido de evidenciar e até mesmo destacar aspectos ligados à frequência e à gravidade de acidentes no trânsito.

MATERIAL E MÉTODOS

O banco de dados analisado neste projeto é público e fora obtido diretamente na página do projeto *Kaggle* (2017). Originalmente, os dados foram coletados pela polícia local do Reino Unido. Como não há a obrigatoriedade dos cidadãos de informar a oficiais sobre acidentes, menos em situações específicas, pode-se afirmar que não há a representação do espectro total dos acidentes ocorridos no país. Este foi o preâmbulo desta pesquisa, focada na análise e na interpretação deste conjunto de informações adquiridas no período entre janeiro de 2005 até dezembro de 2015.

Neste projeto, foi realizado um estudo de casos sobre uma base de dados consistente e rica em informações como condições da rodovia, de luminosidade, do clima, tipo de veículo envolvido, severidade do acidente, número de acidentados, de acidentes, fora a localização e a data do ocorrido. A análise foi feita através da obtenção de relações entre informações contidas no banco de dados em conjunto de diversas técnicas de física estatística, algoritmos de aprendizado de máquina e de softwares livres relacionados à plataforma *Jupyter* e às bibliotecas do *Python*, além de códigos de desenvolvimento próprio.

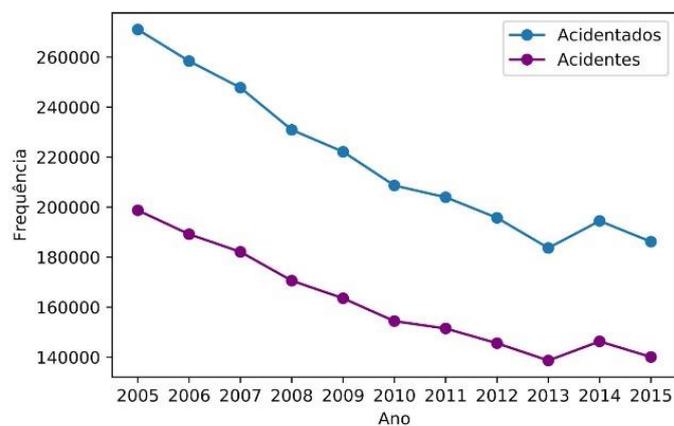
RESULTADOS E DISCUSSÃO

É perceptível que, para cada acidente, há peculiaridades e condições específicas para a sua ocorrência. Entretanto, é preciso sondar e identificar a associação e correlação entre estes de forma estatística, a fim de evidenciar comportamentos e paradigmas dentre essas variáveis. Para tal, realizou-se um processo de visualização dos dados, de forma a gerar histogramas e gráficos de

dispersão, estes feitos em relação aos pares de variáveis, já aqueles feitos em relação a uma única variável.

Ao observar a Figura 1, fica evidente que, com o decorrer dos anos há uma tendência decrescente no número de acidentados, que pode decorrer de uma série de fatores, dentre estes: recessões econômicas, tal qual a de 2008, que ocorreu durante o período analisado; desenvolvimentos e melhorias na infraestrutura e na tecnologia das estradas e dos veículos; criação de centros voltados ao cuidado de traumas. Pode-se perceber que isto ocorre até 2013, enquanto que essa tendência é perdida no período entre 2013 e 2014, porém aparenta ressurgir a partir de 2014 (DEPARTMENT FOR TRANSPORT, 2014, p. 7)

Figura 1 - Frequência de acidentados e acidentes por ano, no período de 2005 a 2015

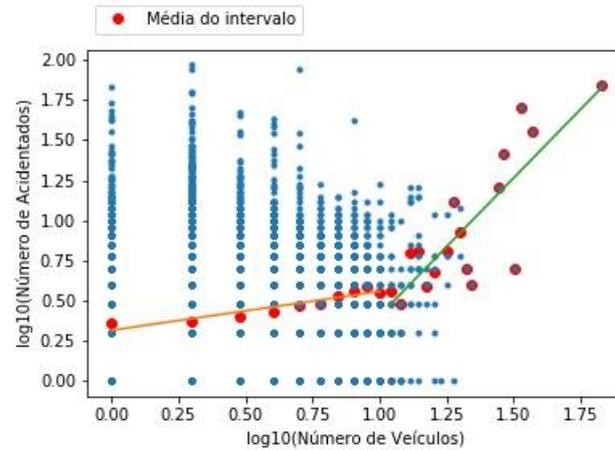


Fonte: Próprio Autor

O aumento repentino no número de acidente e acidentados no ano de 2014 foi percebido pelo governo, que, desde 1997 não havia vislumbrado um aumento em relação tanto ao número de acidentes quanto com os de acidentados, tanto que foram discutidos os motivos que causaram este incremento de acidentes e acidentados em seu relatório anual sobre tráfego de automóveis no ano de 2014. Esse efeito, foi associado a fatores que contribuem para a ocorrência de acidentes. Dentre esses, os que mais se destacam são relacionados ao comportamento do motorista, seus erros e reações, a inexperiência e a imprudência. Como todos os fatores nomeados estão ligados a falhas humanas, não expressa diretamente alguma má manutenção das vias e suas condições (DEPARTMENT FOR TRANSPORT, 2014, p. 3-5).

Analisando os dados em escala logarítmica encontram-se comportamentos interessantes, assim como o da Figura 2, onde é plotado o logaritmo na base dez do número de veículos em função do logaritmo na base dez do número de acidentados em que é possível observar uma transição de fase entre dois regimes alométricos utilizando a média dos intervalos como base. Nota-se que o número dez veículos envolvidos num acidente mostra-se um valor crítico, foi feito o ajuste dos dados por meio de uma regressão linear para os períodos de [0:1] e [1:1,75], obteve-se a inclinação das duas retas de, respectivamente 0,243 e 1,72, que são comportamentos bastante distintos.

Figura 2 - Alometria entre número de veículos e número de acidentados

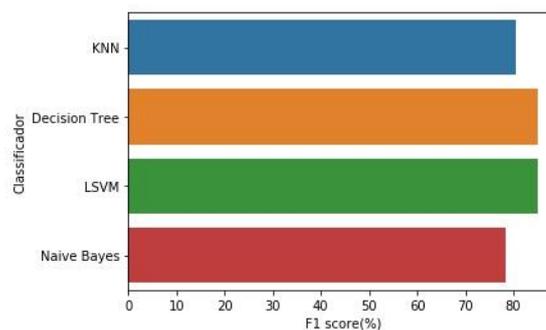


Fonte: Próprio Autor

Numa tentativa de desenvolver um modelo de classificação de alguma característica da base de dados, destacou-se a severidade do acidente, que é dividida em fatal, sério e leve e que parece ser influenciada pelas especificidades de cada acidente, tais quais a localização do acidente, o número de veículos envolvidos, o número de acidentados, a presença de sinalização na rodovia, dentre outros.

Este modelo foi criado baseando-se nas seguintes técnicas de aprendizado de máquina supervisionado: *K-Nearest Neighbors (KNN)*, que segundo Allibhai (2018), realiza um agrupamento de um certo dado observando os 'k' pontos mais próximos deste, atribuindo-lhe uma classificação com base na maioria destes pontos; *Decision Tree*, de acordo com Navlani (2018), segmenta os dados com base em determinados fatores da própria base de dados, criando o que são chamados de nós de decisão até que se todos estes fatores sejam utilizados, resultando nas classificações; *Linear Support Vector Machine (LSVM)*, conforme VanderPlas (2016) é um método de classificação discriminativa que busca traçar uma fronteira mais favorável entre o conjunto de dados; *Naive Bayes* calcula a probabilidade de dada informação pertencer a determinada classe usando probabilidade condicional, assumindo que não há relação entre as características da base de dados, o que é incorreto, mas permite a simplificação do cálculo em grande escala (MÜLLER, 2018).

Figura 3 - Métrica de avaliação dos classificadores



Fonte: Próprio Autor

A fim de avaliar a performance dos métodos utilizados para classificar os acidentes de acordo com sua severidade, pode-se observar na Figura 3 utilizando a métrica *F1 score*, que se baseia na acurácia e revocação do modelo e leva em conta a distribuição desigual das classes de severidade do acidente, visto que há um número muito maior de acidentes leves do que fatais. Destacaram-se os métodos *Decision Tree* e *Linear Support Vector Machine*, que obtiveram respectivamente um *F1 score* de 85,1531 e 85,1536, o que representa um desempenho relativamente bom (JOSHI, 2016).

CONCLUSÕES

A partir deste trabalho foi possível realizar uma investigação de dados de acidentes no Reino Unido no período de 2005 a 2015, evidenciando uma diminuição no número de acidentes e no de acidentados. Nesse intervalo, a tendência negativa pode representar uma melhoria nas providências governamentais em relação ao que abrange o tráfego no país, com uma leve flutuação crescente nesse número do ano de 2013 para 2014. Pode-se observar também um comportamento de dois regimes alométricos entre o número de veículos e o número de acidentados, com destaque para o número de dez veículos, que se mostrou um número crítico. Além disso verificou-se que é factível desenvolver um modelo de classificação em relação à severidade de cada acidente com um desempenho satisfatório.

AGRADECIMENTOS

Agradeço à UTFPR pela bolsa de iniciação científica, que contribuiu para o desenvolvimento desse trabalho e ao professor e orientador Fernando José Antônio pela oportunidade dada.

REFERÊNCIAS

ALLIBHAI, E. **Building a k-Nearest-Neighbors (k-NN) Model with Scikit-learn**. Disponível em: <<https://towardsdatascience.com/building-a-k-nearest-neighbors-k-nn-model-with-scikit-learn-51209555453a>>. Acesso em: 1 ago. 2019.

CLARKE, D. D.; FORSYTH, R., WRIGHT, R. Machine learning in road accident research: decision trees describing road accidents during cross-flow turns. *Ergonomics*, Reino Unido, v. 41, n. 7, p. 1060-1079, 1998.

DEPARTMENT FOR TRANSPORT. **Contributory factors to reported road accidents 2014**. Disponível em: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/463043/rrcgb2014-02.pdf>. Acesso em: 23 jul. 2019.

DEPARTMENT FOR TRANSPORT. **Factors affecting reported road casualties**. Disponível em: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/556406/rrcgb2015-02.pdf>. Acesso em: 16 jun. 2019.

DEPARTMENT FOR TRANSPORT. **Reported Road Casualties in Great Britain: Main Results 2013.** Disponível em: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/324580/rrcgb-main-results-2013.pdf>. Acesso em: 23 jul. 2019.

JOSHI, R. **Accuracy, Precision, Recall & F1 Score: Interpretation of Performance Measures.** Disponível em: <<https://blog.exsilio.com/all/accuracy-precision-recall-f1-score-interpretation-of-performance-measures/>>. Acesso em: 1 ago. 2019.

KAGGLE. **UK Accidents 2005-2015.** Disponível em: <<https://www.kaggle.com/silicon99/dft-accident-data>>. Acesso em: 5 mar. 2018.

MÜLLER, M. **Naive Bayes Classification With Sklearn.** Disponível em: <<https://blog.sicara.com/naive-bayes-classifier-sklearn-python-example-tips-42d100429e44>>. Acesso em: 1 ago. 2019.

NAVLANI, A. **Decision Tree Classification in Python.** Disponível em: <<https://www.datacamp.com/community/tutorials/decision-tree-classification-python>>. Acesso em: 1 ago. 2019.

VANDERPLAS, J. **In-Depth: Support Vector Machines.** Disponível em: <<https://jakevdp.github.io/PythonDataScienceHandbook/05.07-support-vector-machines.html>>. Acesso em: 1 ago. 2019.

WORLD HEALTH ORGANIZATION. **Global status report on road safety.** Disponível em: <https://www.who.int/violence_injury_prevention/road_safety_status/2015/GSRRS2015_Summary_EN_final2.pdf?ua=1>. Acesso em: 22 jul. 2019.

WORLD HEALTH ORGANIZATION. **The top 10 causes of death.** Disponível em: <<https://www.who.int/en/news-room/fact-sheets/detail/the-top-10-causes-of-death>>. Acesso em: 22 jul. 2019.