

MODELAGEM DE COMPORTAMENTO DE MOTORISTAS POR MEIO DE DADOS GERADOS POR TELEMETRIA

Viviane Cardosina Cordeiro¹ (IC), Dr. Bruno Tardiole Kuehne (PQ)¹

¹Universidade Federal de Itajubá

Palavras-chave: Acidentes de Trânsito. Base de Dados. Comportamento de Condução. Perfil de Motorista. Rede CAN.

Introdução

O comportamento de motoristas, quando estão conduzindo seus veículos, é um dos fatores determinantes para a segurança no trânsito. Segundo Rozestraten (1988), o principal motivo de acidentes é a conduta dos usuários das vias, que não é compatível com a segurança no tráfego. Esse comportamento pode diminuir a quantidade de acidentes ou contribuir para que os números sejam ainda maiores. Muitos custos são gerados ao poder público em decorrência desses acidentes, sejam eles custos diretos (como serviços médico-hospitalares) ou indiretos (como perda de produção e produtividade).

O ato de dirigir um veículo é uma tarefa comum e também complexa. Pessoas saem de suas casas diariamente e destinam-se para um passeio, para realizar uma determinada tarefa ou para ir até o local de trabalho. E, apesar de ser uma tarefa corriqueira, envolve muita atenção e inúmeras subtarefas: mudar de marcha, sinalizar uma troca de faixa, o uso do pedal de freio ou do pedal de aceleração, dentre outras. Durante o percurso o motorista é constantemente afetado por acontecimentos internos e externos ao veículo. Múltiplas decisões precisam ser tomadas baseado no ambiente de tráfego em que o veículo se encontra, nas condições climáticas e no estado físico e mental do condutor. Todos os motoristas estão sujeitos a essas circunstâncias e cada um possui um comportamento de condução que os difere dos demais. Seja pelo tempo em que demoram para girar o volante, como utilizam o pedal de freio ou o pedal de aceleração, cada indivíduo tem as suas diferenças e são essas características que possibilitam identificar uma impressão digital do motorista.

A rede veicular é responsável pela comunicação eletrônica entre os diferentes dispositivos presentes nos automóveis. Os protocolos de comunicação possuem características e formações próprias à sua aplicação dependendo da taxa e do tipo de transmissão, por exemplo. Considerando todas as informações que o monitoramento dos sistemas veiculares conseguem fornecer, o reconhecimento do comportamento dos

motoristas como seguro ou inseguro, baseado em padrões de condução e alimentando as informações de como se portam, pode ajudar a promover uma condução segura, reduzir o número de acidentes e contribuir para a segurança social (DONG, WU, ZHANG, 2016). Essas informações também possibilitam customizar o veículo de acordo com seu motorista. Seguradoras podem diminuir o valor de seguros, caso o comportamento seja classificado como seguro, e ainda, sistemas avançados podem ser implementados a fim de fornecer assistência ao motorista e melhorar o desempenho de cada condutor.

Analisar e entender o modo de condução é uma tarefa necessária a muitos cientistas e pesquisadores. Projetistas e engenheiros buscam entender esse comportamento para que possam tomar as melhores decisões e garantir que as soluções de assistência funcionem adequadamente em sistemas de tráfego dinâmico (ZARDOSHT, 2016). O comportamento de dirigir é complicado e difícil de analisar e, apesar de não ser algo novo, não há estudos que identifiquem uma técnica ótima em aplicações reais.

Essa pesquisa busca estudar uma função que permita aprimorar e ampliar os conhecimentos sobre o comportamento do motorista e assim identificar o perfil do condutor. Para isso deve-se passar pela análise de quais dados são relevantes para o estudo e, posteriormente, determinar se o comportamento de determinado motorista é seguro ou inseguro.

No primeiro momento, fez-se detalhada investigação das diferentes causas que podem levar a ocorrência de acidentes durante a condução de veículos. Também dedicou-se tempo ao estudo de trabalhos já publicados sobre a modelagem de comportamento de motoristas e, em seguida, foi realizado o tratamento dos dados coletados da rede CAN dos caminhões monitorados para este trabalho.

Metodologia

Coletar dados sempre envolve diversos fatores, internos e externos, que podem resultar em interferência nos

resultados obtidos. Nos sistemas de telemetria isso não é diferente e para que possam ser melhor analisados é necessário que, primeiramente, sejam tratados.

Os dados aqui estudados foram extraídos através de um módulo de conectividade instalado em cada um dos caminhões. Eles foram coletados em um espaço de tempo determinado durante viagens reais, em ambientes não controlados e de motoristas diferentes. Esses dados correspondem a informações já existentes na rede CAN veicular e estão constantemente sendo enviados por diferentes módulos eletrônicos nos veículos. No processo de extração, os dados foram primeiro encaminhados para uma nuvem da Volkswagen Caminhões e Ônibus (VWCO) e posteriormente foram acessados de uma plataforma *web*. A nuvem se encontra em servidores da Amazon AWS e, para cada dia de coleta, os dados são extraídos e compilados em uma planilha Excel por um engenheiro de dados da VWCO.

Os automóveis utilizados nesta pesquisa foram caminhões dos tipos Constellation, que abrange a categoria de médios e pesados, e Delivery, que corresponde a faixa de modelos com peso bruto total de 3.5 a 13 toneladas.

Foi escolhida a linguagem de programação Python, juntamente com a ferramenta Jupyter Notebook, para realização da análise. O Jupyter Notebook é um software gratuito, com visual simples, de fácil utilização e a instalação da plataforma Anaconda proporciona sua utilização em uma máquina local, utilizando um navegador que esteja disponível. Os dados aqui analisados foram disponibilizados em planilhas Excel com extensão *xlsx*, sendo que cada arquivo corresponde a um dia específico de viagem. Inicialmente, esse trabalho se concentra em analisar os dados das viagens de um único motorista, de modo que o mesmo procedimento pode ser aplicado posteriormente aos demais.

Iniciou-se o tratamento dos dados excluindo a primeira linha de todos os arquivos. A linha correspondia ao título da tabela com informações sobre a data em que os dados haviam sido coletados. Essa ação não prejudica a análise, pois a mesma se encontra disponível na primeira coluna para cada um dos dados. Esse passo foi necessário para possibilitar a concatenação das tabelas e facilitar a observação dos dados. Utilizou-se a função *glob()* e comandos da biblioteca *Openpyxl* para especificar o caminho até os arquivos e para exclusão da linha, respectivamente. Para carregar e concatenar os arquivos foram utilizadas funções da biblioteca *Pandas*. Após a concatenação dos dados, o novo arquivo passou a ter uma primeira linha correspondente ao título de cada coluna e mais de 677 mil linhas de informações.

Optou-se por utilizar as variáveis que caracterizavam a

economia de combustível, a velocidade e a utilização do pedal de freio e de aceleração. As colunas correspondentes a essas informações foram filtradas dos dados iniciais e passaram a compor um novo arquivo.

A figura 01, correspondente às 3 primeiras linhas do novo arquivo, mostra que alguns dados estavam acompanhados de suas respectivas unidades. Para que elas fossem removidas criou-se uma função que recebe o *Data Frame (DF) pandas* (o arquivo de dados), uma coluna e qual dado que deve ser deletado. Para que essa informação não fosse perdida, a mesma foi adicionada ao título de sua respectiva coluna.

	Data	UTC Time	Economia de combustível	Velocidade Real	Velocidade Do Motor	Pedal De Freio	Posição Do Pedal De Aceleração
0	2021-02-01 06:41:40	06:41:40	9 [km/l]	0 [km/h]	756 [rpm]	0 [bin]	0 [%]
1	2021-02-01 06:41:42	06:41:42	9 [km/l]	0 [km/h]	752 [rpm]	0 [bin]	0 [%]
2	2021-02-01 06:41:44	06:41:44	13 [km/l]	0 [km/h]	754 [rpm]	0 [bin]	0 [%]

Figura 01 - Primeiras linhas de dados

Ao buscar informações sobre os dados percebe-se que, mesmo após a remoção das unidades, eles ainda são do tipo *object* e para uma melhor interpretação precisam ser convertidos para dados do tipo *float64*. Para isso foi utilizado o método *astype()* que permite lançar um objeto *Pandas* para um tipo específico. Como mencionado anteriormente, os dados podem ser corrompidos durante a sua extração e a biblioteca *Missingno* permite visualizar a distribuição de valores faltantes (*Not a Number - NaN*) no *data frame*. Na figura 02 é possível observar linhas brancas, que correspondem aos valores ausentes, e linhas pretas, representando aos dados presentes no DF.

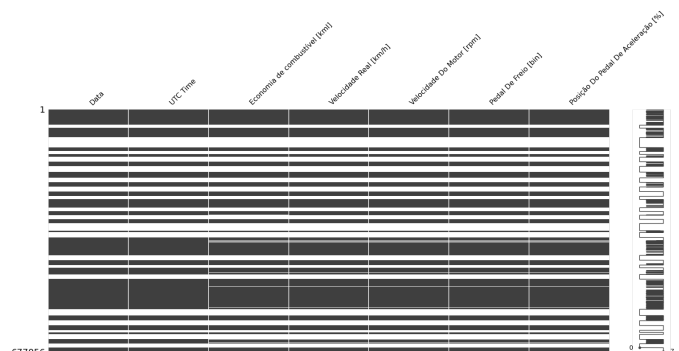


Figura 02 - Dados faltantes

O emprego do método *Pandas dropna()* possibilita a exclusão dessas linhas. Por fim, verificou-se a existência de linhas com dados duplicados que foram eliminados

aplicando o método `drop_duplicates()`. O novo DF possui 403362 linhas de dados conforme apresentado no canto inferior esquerdo da figura 03.

	Data	UTC Time	Economia de combustível [kml]	Velocidade Real [km/h]	Velocidade Do Motor [rpm]	Pedal De Freio [bin]	Posição Do Pedal De Aceleração [%]
...
403357	2020-12-30	20:27:47	0	9	1351	0	1
403358	2020-12-30	20:27:49	5	5	764	1	0
403359	2020-12-30	20:27:51	16387	1	754	0	0
403360	2020-12-30	20:27:53	3	0	748	1	0
403361	2020-12-30	20:27:55	3	0	749	1	0

403362 rows x 7 columns

Figura 03 - Últimas linhas de dados

Resultados e discussão

Na primeira etapa deste trabalho foi realizada a investigação do estado da arte das diferentes causas para a ocorrência de acidentes de trânsito. Foi constatado que a falha humana na percepção de perigo é o principal fator que pode levar a um acidente. Fatores como excesso de velocidade, desatenção, uso de drogas lícitas e a pouca experiência de condução também foram citados. Desse modo, fica claro a importância de se estudar soluções que auxiliem na redução desses números.

Ao realizar o agrupamento dos dados era esperado que alguns deles estivessem corrompidos e estes foram retirados da base de dados para que não influenciassem negativamente a análise. Após o devido tratamento, foi gerado o histograma da figura 04 que permite observar o comportamento do motorista com relação à velocidade atingida pelo caminhão durante as viagens. Vale acrescentar que os dados começaram a ser coletados antes da saída do automóvel, ou seja, há um tempo de espera anterior à partida que também está presente nas informações. Em todos os gráficos os valores na vertical correspondem ao número de dados.

É possível observar duas faixas de velocidades nas quais o motorista esteve a maior parte do tempo, sendo elas de 0 a 10 km/h e de 70 km/h a 90 km/h, aproximadamente. O próximo gráfico, figura 05, mostra como ocorreu a utilização do pedal de freio durante as viagens. Foi definido um valor 0 quando o pedal não estava sendo utilizado e um valor 1 quando o pedal era acionado. Fica claro que na maior parte do tempo o motorista não fez uso do pedal de freio. Com relação ao pedal de

aceleração, a figura 06 mostra o caminhão na maior parte do tempo em uma faixa de 0 a 10% do pedal aceleração pressionado, aproximadamente. Percebe-se ainda, que o motorista chega a utilizar 100% do pedal de aceleração em partes da viagem.

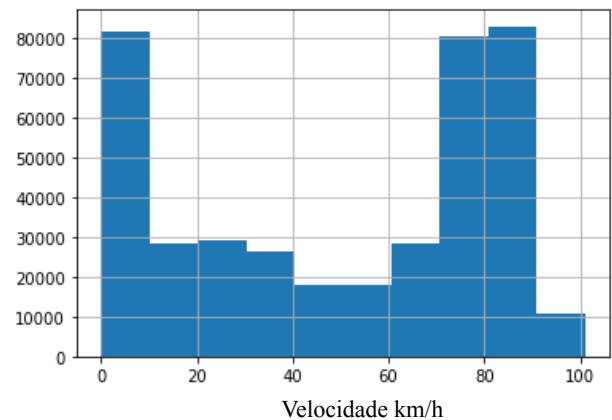


Figura 04 - Gráfico das velocidades atingidas durante as viagens.

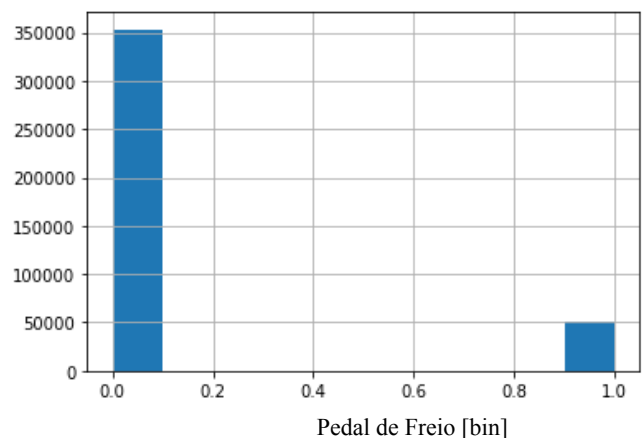


Figura 05 - Gráfico da utilização do pedal de freio nas viagens

Com base nos resultados aqui apresentados, o próximo passo da pesquisa será dedicado a encontrar os parâmetros ótimos a serem empregados nos procedimentos de análise de dados, assim como a implementação de técnicas de inteligência artificial que consigam encontrar o perfil do motorista no menor tempo possível e com a maior acurácia.

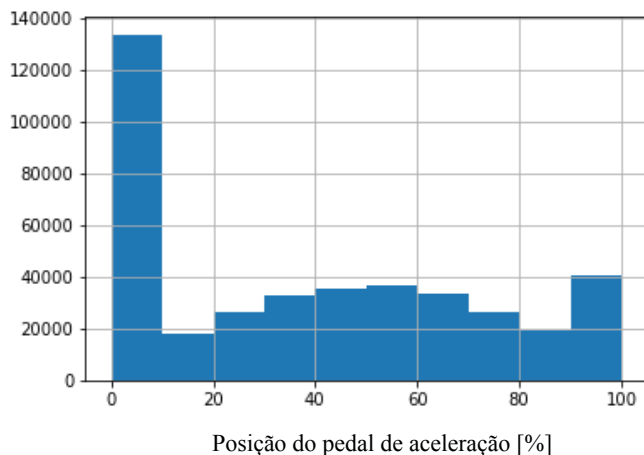


Figura 06 - Gráfico da utilização do pedal de aceleração nas viagens

Conclusões

Identificar características de condutores e utilizá-las para obter maior controle automotivo é uma atividade informativa e bastante ampla. Analisar o comportamento de motoristas diferentes é uma atividade complexa, visto que o indivíduo pode responder de diferentes formas em diferentes situações de condução. Para mais, diversas implicações de tráfego e condições climáticas também podem afetar as características do condutor.

No presente trabalho, estudou-se a ligação intrínseca entre as características do comportamento de dirigir e suas implicações para a segurança no trânsito. Baseado na literatura, foi especificado as características do comportamento do motorista que compõem a base de dados desta pesquisa, sendo elas: economia de combustível, a velocidade e a utilização do pedal de freio e de aceleração. Os dados irrealis foram retirados da análise, os demais receberam o devido tratamento e a continuidade da pesquisa se destina a encontrar os parâmetros ótimos para a implementação de técnicas de inteligência artificial e aprendizado de máquina utilizando os dados obtidos a fim de identificar de maneira eficiente um perfil para o motorista. Como as classificações existentes são ainda grosseiras, deve-se refinar ainda mais os novos estudos.

Agradecimento

Primeiramente agradeço a Deus pelo presente da vida. Ao meu orientador e também professor Dr. Bruno Tardiole Kuehne, pelas instruções, apoio e toda compreensão ao longo do desenvolvimento da pesquisa. Ao professor Dr. Otávio de Souza Martins Gomes e ao engenheiro da Volkswagen e aluno de mestrado, Lucas Gomes de Almeida, pela ajuda e empenho com seu

conhecimento.

Agradeço também a Coordenação de Pesquisa e Pós-Graduação da UNIFEI e ao CNPq, essenciais no meu processo de formação acadêmica e por fomentar o interesse de discentes à pesquisa por meio de Iniciação Científica.

Referências

ROZESTRATEN, R. Psicologia do trânsito: conceitos e processos básicos. São Paulo: EPU; Edusp. p. 154. 1988

ZARDOSHT, M. *Identifying Individual Driver Behaviour Using In-Vehicle CAN-bus Signals of Pre-Turning Maneuvers*. London, Electronic Thesis and Dissertation Repository. 4160, out. 2016.

DONG, Y., WU, M., ZHANG, S. A Novel Model-Based Driving Behavior Recognition System Using Motion Sensors. *Sensors*, China, p.16, 20 out. 2016