

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: Análise de dados, Clusters, K-Means, PCA, Perfil de Motoristas.

Introdução

O comportamento dos motoristas ao conduzir veículos é um fator determinante para a segurança nas estradas. Conforme ressaltado por Rozestraten (1988), a principal causa de acidentes de trânsito está intrinsecamente ligada aos comportamentos dos usuários das vias, frequentemente incompatíveis com as normas de segurança. Esse comportamento pode influenciar tanto na diminuição quanto no aumento do número de acidentes.

Embora dirigir um veículo seja uma atividade comum, ela é igualmente complexa. Requer constante atenção e envolve a realização de múltiplas tarefas simultaneamente. Durante o percurso, os motoristas estão sujeitos a uma série de fatores externos e internos, como as condições do tráfego, a meteorologia e seu próprio estado físico e mental. Cada motorista responde de maneira única a essas circunstâncias, apresentando diferenças na velocidade de reação e nas técnicas de condução. Essas variações tornam possível a identificação de um "perfil" individual do motorista, que, de certa forma, se assemelha a uma impressão digital.

O uso de sistemas de monitoramento veicular pode contribuir substancialmente para aprimorar a segurança no trânsito por meio da caracterização do comportamento dos condutores. Informações detalhadas sobre o uso dos pedais de freio e acelerador podem ser empregadas para construir esse perfil. Além disso, sistemas avançados podem ser implementados para oferecer assistência ao motorista e otimizar o desempenho de cada um.

A análise e compreensão do comportamento de condução são tarefas cruciais tanto para cientistas quanto para pesquisadores. Projetistas e engenheiros buscam compreender esse comportamento a fim de tomar decisões embasadas e garantir o funcionamento adequado de sistemas de assistência em ambientes de tráfego dinâmico (ZARDOSHT, 2016). Contudo, é importante salientar que o comportamento de condução é intrincado e desafiador de analisar, não havendo uma técnica universalmente ideal para aplicações no mundo

real. Este estudo tem como objetivo desenvolver uma abordagem que aprimore a compreensão do comportamento do motorista e permita a identificação de perfis de condutores, com foco nas diferenças entre a condução segura e a condução insegura. Para tanto, envolve a extração de dados da rede, a identificação dos dados mais relevantes e a caracterização do comportamento de diferentes motoristas.

Metodologia

Ao realizar a coleta de dados, diversos fatores, tanto internos quanto externos, podem interferir nos resultados. Isso também se aplica a sistemas de telemetria, onde é crucial pré-processar os dados antes de analisá-los.

Nessa pesquisa, optou-se pela utilização da linguagem de programação Python e a ferramenta Colab, que é uma plataforma de computação em nuvem de acesso gratuito, oferecida pela Google, que permite aos usuários escrever, executar e compartilhar códigos.

Os dados analisados neste estudo foram obtidos por meio de um módulo de conectividade instalado em um caminhão do tipo Delivery, que compreende modelos com pesos brutos totais variando de 3,5 a 13 toneladas. Esses dados foram coletados durante viagens reais, em ambientes não controlados em todo o estado de São Paulo e por um único motorista. Eles correspondem a informações que já estão disponíveis na rede CAN (Controller Area Network) do veículo e são continuamente transmitidos por diversos módulos eletrônicos nos veículos.

No processo de extração, os dados foram inicialmente enviados para uma plataforma de nuvem hospedada na infraestrutura da Amazon AWS, especificamente na nuvem da Volkswagen Caminhões e Ônibus (VWCO). Posteriormente, esses dados foram acessados por meio de uma plataforma web. Para cada dia de coleta, um engenheiro de dados da VWCO extraiu e compilou os dados em uma planilha Excel.

Os dados foram concatenados em uma única tabela com aproximadamente 677 mil linhas de dados. As variáveis consideradas para o estudo correspondem à velocidade,

em Km/h, ao uso do pedal de freio, em binário (com 1 para pressionado e 0 para não pressionado), e a porcentagem de utilização do pedal de aceleração.

Os dados foram tratados e, em seguida, iniciou-se a implementação de técnicas de inteligência artificial e aprendizado de máquina na tentativa de encontrar aquela que tem o melhor desempenho no agrupamento dessas informações.

Como os dados são não supervisionados, ou seja, não possuem rótulos ou classes pré-definidas, e estamos analisando 3 diferentes variáveis, velocidade, pedal de freio e posição do pedal de aceleração, primeiro utilizou-se uma abordagem de redução de dimensionalidade e depois foram feitas as tentativas de agrupamento dos dados.

A redução da dimensionalidade proporciona a redução do número de características mantendo as informações mais importantes. Aqui, utilizou-se a técnica PCA (Análise de Componentes Principais) com os parâmetros $n_components = 2$ e $random_state = 42$, para esse fim. A figura 01, obtida com a função *scatter* da biblioteca *Matplotlib*, mostra o gráfico de dispersão após a redução de dimensionalidade.

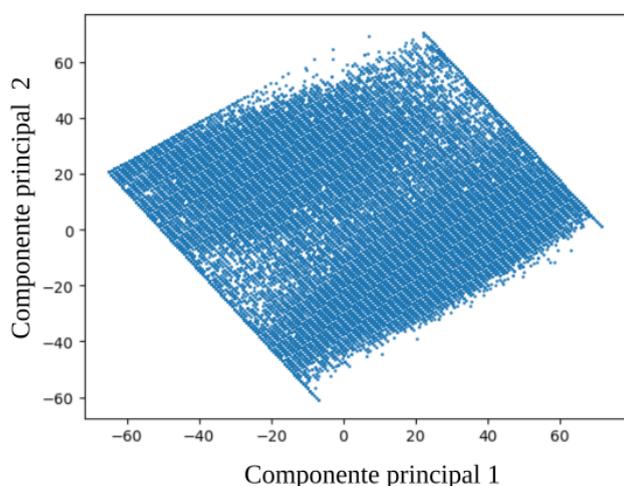


Figura 01 - Gráfico de dispersão do PCA

Em sequência iniciou-se os testes com os algoritmos de *clustering* (agrupamento): *Birch*, *Gaussian Mixture*, *K-means*, *MeanShift* e *OPTICS*. Esses algoritmos visam dividir os dados em *clusters* com base em suas características, de modo que itens semelhantes estejam no mesmo grupo.

As métricas *Calinski-Harabasz Score*, que mede a relação entre a dispersão entre os *clusters* (variância entre *clusters*) e a dispersão dentro dos *clusters* (variância intracluster), *Davies-Bouldin Score*, que mede a média das razões das distâncias médias entre pontos de dados dentro de um *clusters* e pontos de dados

em outros *clusters* e a *Silhouette Score*, que mede o quão similar um ponto de dados é ao seu próprio *cluster* (coesão) em comparação com outros *clusters* (separação), foram utilizadas para trazer uma melhor compreensão dos dados.

Tanto os testes com os algoritmos de agrupamento quanto o cálculo das métricas foram realizados com e sem a utilização de PCA nos dados.

Resultados e discussão

A primeira etapa deste trabalho envolveu uma análise do estado atual das causas de acidentes de trânsito. Constatou-se que a falha humana na percepção de riscos é uma das principais causas de acidentes, incluindo fatores como excesso de velocidade, distração, uso de substâncias lícitas e falta de experiência na condução. Isso enfatiza a necessidade de explorar soluções para reduzir esses incidentes.

Na segunda parte da pesquisa, tratamos os dados provenientes da telemetria da rede CAN de caminhões. Durante o processo de agrupamento dos dados, identificamos informações corrompidas que foram excluídas para evitar distorções na análise. Após o tratamento adequado, geramos um histograma (Figura 02) para observar o comportamento dos motoristas em relação à velocidade atingida pelos caminhões durante as viagens. Nas figuras 02, 03 e 04, os valores na vertical correspondem ao número de leituras, e cada leitura foi feita em um intervalo de 2 segundos, aproximadamente. Vale ressaltar que a coleta de dados começou antes da partida do veículo, incluindo um período de espera anterior ao início da viagem.

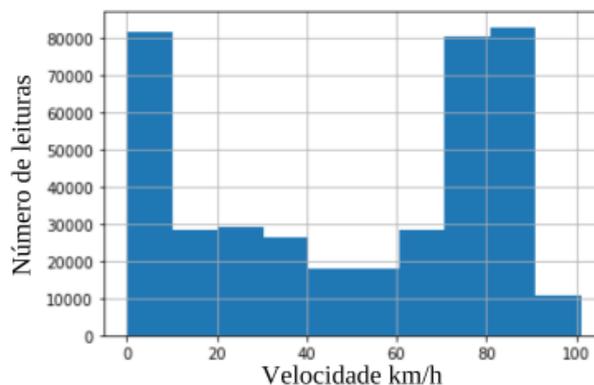


Figura 02 - Gráfico das velocidades atingidas.

Podemos notar duas faixas de velocidade em que o motorista passou a maior parte do tempo: uma delas entre 0 a 10 km/h e a outra entre 70 km/h e 90 km/h, aproximadamente. O gráfico subsequente ilustra o uso do pedal de freio durante as viagens. Atribuímos o valor

0 quando o pedal não estava sendo pressionado e o valor 1 quando estava em uso.

Na Figura 03, é evidente que o motorista não acionou o pedal de freio na maior parte do tempo. Já em relação ao pedal de aceleração, a Figura 04 revela que o caminhão permaneceu, em grande parte do tempo, com o pedal de aceleração pressionado em uma faixa de cerca de 0 a 10%. Nota-se também que, em certos trechos da viagem, o motorista chegou a utilizar 100% do pedal de aceleração.

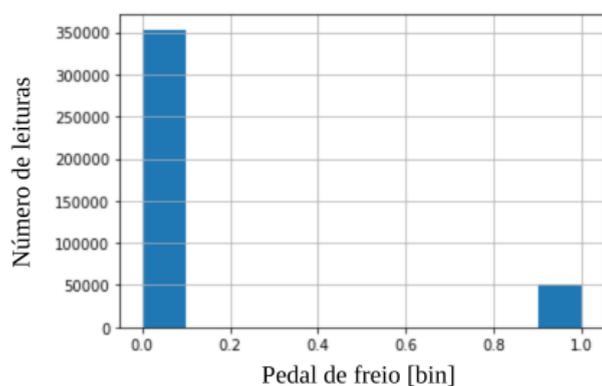


Figura 03 - Gráfico da utilização do pedal de freio nas viagens

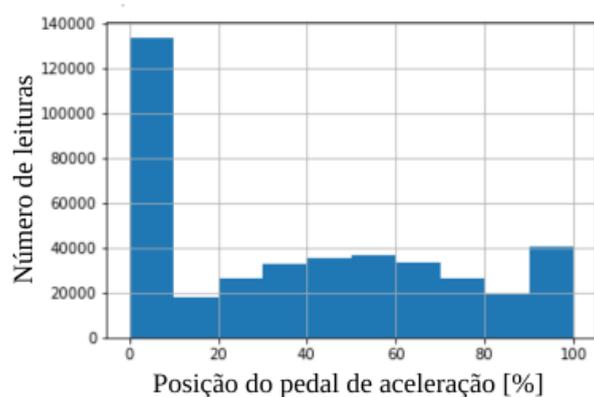


Figura 04 - Gráfico da utilização do pedal de aceleração nas viagens

A visualização dos gráficos de dispersão, após a utilização dos algoritmos de agrupamento, indica que o *K-Means* apresenta o melhor desempenho para o grupo de dados fornecido, e isso acontece com e sem a utilização da PCA. As figuras 05 e 06 correspondem ao gráfico de dispersão do algoritmo *K-Means* com e sem a técnica PCA, respectivamente.

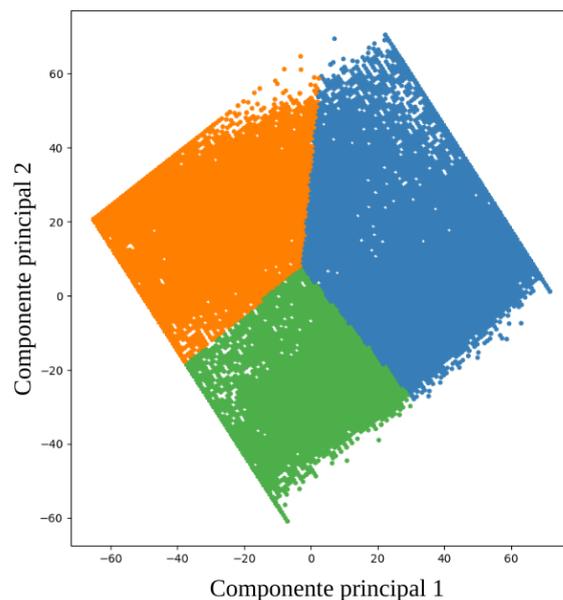


Figura 05 - Gráfico de dispersão do algoritmo *K-Means* com a PCA

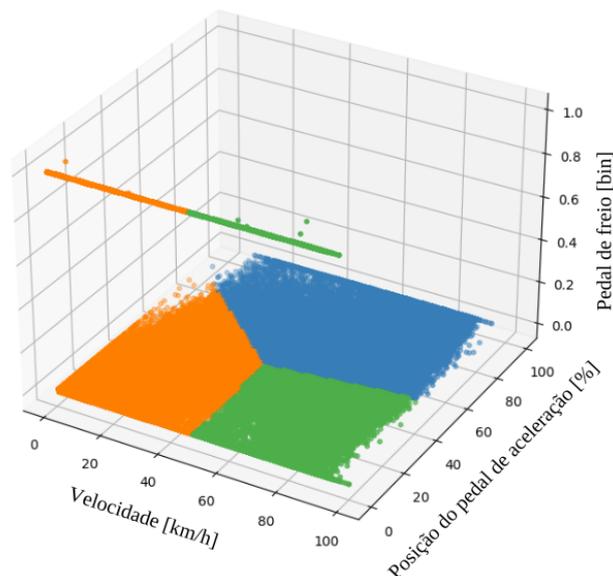


Figura 06 - Gráfico de dispersão do algoritmo *K-Means* sem a PCA

Para uma melhor compreensão dos resultados obtidos, têm-se na tabela 01 os valores de cada uma das métricas empregadas para todos os algoritmos estudados, sendo os valores A e B com e sem PCA, respectivamente. Para a métrica *Calinski Harabasz*, quanto maior o valor da pontuação, melhor a qualidade dos *clusters*, logo, observando a tabela 01, o algoritmo *K-Means* obteve o melhor desempenho. Já para a métrica *Davies Bouldin*, quanto menor o valor, melhor a qualidade dos *clusters*.

Nesse caso, o algoritmo *K-means* demonstrou a melhor eficiência. E a última métrica aplicada, *Silhouette*, na qual os valores mais próximos de +1 indicam uma melhor qualidade dos *clusters*, mostra que o algoritmo OPTICS teve o melhor desempenho, seguido de K-Means em segundo lugar e *Mean Shirt* em terceiro lugar.

Algoritmo	<i>Calinski Harabasz</i>	<i>Davies Bouldin</i>	<i>Silhouette</i>
K-means	A:117980,1 B:97781,30	A:0,7411 B:0,8870	A:0,4692 B:0,4602
Mean Shirt	A:90673,8 B:90673,82	A:0,8888 B:0,8888	A:0,4502 B:0,4502
OPTICS	A:196,4 B:196,47	A:1,3139 B:1,3139	A:0,7954 B:0,7954
Birch	A:64478,0 B:95377,47	A:0,9844 B:0,9033	A:0,3511 B:0,4523
Gaussian Mixture	A:10390,3 B:20776,76	A:1,6196 B:1,0887	A:0,0654 B:0,2273

Tabela 01 -Valor das métricas para cada algoritmo com (A) e sem (B) PCA

Conclusões

Identificar e aplicar as características individuais dos condutores para aprimorar o controle automotivo é uma tarefa abrangente e de relevância significativa. Analisar o comportamento de motoristas distintos é intrinsecamente complexo, devido à variabilidade nas suas reações perante diversas situações de condução, além das influências do tráfego e das condições climáticas.

Neste estudo, explorou-se a relação essencial entre os comportamentos de direção e sua importância na segurança viária. A partir da literatura existente, identificamos as características do comportamento do motorista que formam a base de dados deste projeto: velocidade do veículo e o uso dos pedais de freio e aceleração. Dados inconsistentes foram excluídos da análise, enquanto os demais foram devidamente processados.

Os dados utilizados nesta pesquisa são de um único motorista, coletados durante várias viagens, e o mesmo processo pode ser aplicado a uma quantidade maior de dados e motoristas.

Utilizando-se de técnicas de inteligência artificial e

aprendizado de máquina, foi possível reduzir a dimensão dos dados para apenas duas componentes e, assim, aplicar algoritmos de agrupamento para classificar os dados em três grupos diferentes, levando em consideração as características da velocidade do automóvel, o acionamento ou não do pedal de freio e a porcentagem de utilização do pedal de aceleração em vários pontos das viagens de um determinado motorista. Constatou-se que a utilização da PCA não foi um fator determinante para essa pesquisa e que os resultados podem ser aprimorados mesmo sem sua utilização.

Agradecimentos

Em primeiro lugar, expresse minha profunda gratidão a Deus pelo dom da vida.

Ao meu orientador e também professor, Dr. Bruno Tardiole Kuehne, pelos ensinamentos, apoio e compreensão ao longo do desenvolvimento da pesquisa. Também desejo estender meus agradecimentos ao aluno de mestrado Lucas Gomes de Almeida, pela assistência significativa neste trabalho.

Não posso deixar de reconhecer a importância crucial da Coordenação de Pesquisa e Pós-Graduação da UNIFEI e do CNPq para minha formação acadêmica. Eles desempenharam um papel fundamental no meu desenvolvimento e permanência na universidade.

Referências

BRASIL. Ministério da Infraestrutura. Secretaria Nacional de Trânsito. Registro Nacional de Acidentes e Estatísticas de Trânsito. Brasília, DF: Ministério da Infraestrutura, 2022. Disponível em: <https://www.gov.br/infraestrutura/pt-br/assuntos/transito/arquivos-senatran/docs/renaest>. Acesso em: 09 set. 2022.

MELO, M.; FIGUEIREDO, N. Acidentes de Trânsito: os impactos causados no setor público de saúde e trânsito em Roraima/BR. *Ambiente*, Boa Vista, v. 12, n.3, p. 123-143. 2019.

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