

DESENVOLVIMENTO DE BANCO DE DADOS PARA CONDIÇÕES DE ADERÊNCIA EM PISTAS DE POUSO E DECOLAGEM

Haikel Buganem Busgaib Gonçalves
Francisco Heber Lacerda de Oliveira

Universidade Federal do Ceará
Programa de Pós-Graduação em Engenharia de Transportes

RESUMO

A ciência de dados está se tornando uma atividade presente cada vez mais nas áreas da engenharia, podendo também ser utilizada em atividades de gerenciamento de pavimentos aeroportuários. Dentre os componentes de um Sistema de Gerência de Pavimentos Aeroportuários (SGPA), o banco de dados é um componente crucial para o sucesso de sua implantação. Com isso, o presente trabalho mostra resultados preliminares da confiabilidade do uso do *software* de reconhecimento óptico de caracteres Tesseract OCR para realizar a leitura de relatórios de medição de atrito e macrot textura, obtendo índices de acerto variando entre 44,13% e 99,45%. Além disso, foram estudados os campos de dados dos relatórios com o objetivo de desenvolver uma planilha modelo que deverá ser alimentada por um algoritmo ao final do trabalho.

1. PROPOSTA DE PESQUISA

Gerenciar pavimentos, em seu sentido mais amplo, inclui todas as atividades envolvidas no planejamento, no projeto, na construção, na conservação, na restauração e na avaliação dos pavimentos que fazem parte de uma infraestrutura aeroportuária (ANAC, 2017). De maneira geral, as pistas de pouso e decolagem (PPDs) constituem grande parte do patrimônio dos aeroportos, devendo esses serem observados e mantidos regularmente. No que diz respeito à funcionalidade desses pavimentos, destaca-se a sua capacidade de auxiliar na segurança dos processos de frenagem e aceleração das aeronaves, principalmente quando existem adversidades do clima.

A capacidade de prover uma adequada aderência entre pneu e pavimento está relacionada com o coeficiente de atrito e a macrot textura do pavimento, dados que podem ser obtidos por medições realizadas nas PPDs. Os resultados dessas medições são impressos em relatórios que são encaminhados à Agência Nacional de Aviação Civil (ANAC). No que se diz respeito ao estado em que os relatórios se encontram, observa-se uma variedade de arquivos que contém dados de diferentes naturezas de localidade, data, qualidade de resolução, etc.. Para que se possa trabalhar melhor com esses dados, deve-se pensar em como padronizá-los e organizá-los de forma mais simples, facilitando o processo dos tomadores de decisões de desenvolver informações úteis.

Com isso, o presente trabalho tem o objetivo de propor o uso de ferramentas e conceitos de ciência de dados e da informação para a estruturação de um banco de dados para gerenciamento de pavimentos aeroportuários. Para isso, propõe-se a construção de um algoritmo capaz de ler relatórios padronizados de coeficiente de atrito e de medição de macrot textura de PPDs no formato “.pdf” em imagem com auxílio de um *software* de Reconhecimento Óptico de Caracteres (ROC) e organizar os dados obtidos nessa leitura em planilhas de forma a fornecer informações úteis para tomadas de decisão.

2. REVISÃO BIBLIOGRÁFICA

Dentre os componentes de um Sistema de Gerência de Pavimentos Aeroportuários (SGPA), Haas, Hudson e Zaniewsky (1994) citam o banco de dados como um componente crucial para o sucesso de sua implantação. As informações contidas em um banco de dados contribuem

para o desenvolvimento de modelos de previsão que podem auxiliar os tomadores de decisão tanto em nível de rede como em nível de projeto.

O termo “dado” geralmente é considerado como a base da hierarquia DIKW (*Data, Information, Knowledge and Wisdom*) na qual a informação é entendida como dado processado, conhecimento como informação processada ou interpretada e sabedoria como conhecimento interpretado (HJØRLAND, 2018). Para que os dados possam ser processados até fornecer uma sabedoria, é importante que exista um sistema ou ferramenta que organize e processe esses dados, facilitando as atividades de busca e acesso. Para isso, Cormen *et al.* (2009) recomendam o desenvolvimento de uma estrutura de dados que seja compatível com o *dataset* à qual será aplicada. A aplicação de uma estrutura de dados gera um banco de dados, que é definido por Fedeli, Polloni e Peres (2010) como um conjunto de dados estruturado adequadamente para ser utilizado de forma eficiente por uma diversidade de aplicações.

Fox (2014) cita a importância da definição de um ciclo de vida quando deseja-se trabalhar com dados por facilitar na aquisição, na cura e na preservação da segurança do dado. Zambetti, Pinto e Pezzotta (2019) demonstram que o ciclo de vida de um dado é composto basicamente na criação, aquisição, armazenamento e análise dos dados, podendo os produtos gerados nesse processo serem posteriormente armazenados de forma estratégica. A Figura 1 apresenta um modelo básico de ciclo de vida de dados, que pode ser modificado conforme as necessidades do usuário.

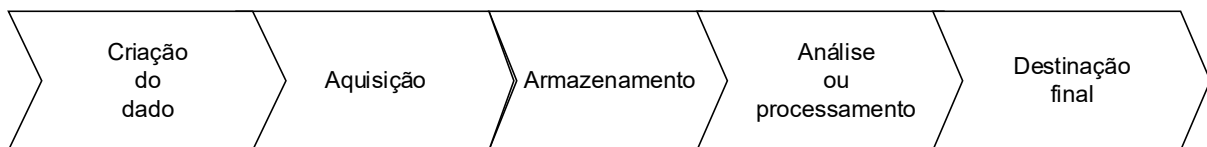


Figura 1: Modelo básico de ciclo de vida de dados

Robby G. *et al.* (2019) reconhecem o uso de algoritmos ROC como uma técnica de converter imagens impressas ou escaneadas em documentos de texto editáveis. Para tal, os autores utilizaram um *software* denominado de Tesseract OCR (*Optical Character Recognition*) por possuir um mecanismo de aprendizado de máquina capaz de ler textos de diferentes fontes e línguas. Para validar a capacidade do ROC, Robby G. *et al.* (2019) calcularam o índice de acurácia obtido nas leituras dos algoritmos, que pode ser determinado pela relação entre o número de acerto de caracteres e o número total de caracteres em uma determinada amostra.

3. MÉTODO DE PESQUISA

Inicialmente foram descritas as necessidades dos usuários determinando os dados que deveriam ser armazenados, as informações que deveriam ser obtidas, quais conhecimentos poderiam ser obtidos ao observar as informações geradas e quais sabedorias poderiam contribuir para modelos de previsões dos dados futuros.

Depois de descritas as necessidades dos usuários, foram modelados organogramas, formatos de tabelas, dicionários e listas que definirão os campos de dados que serão trabalhados para a estruturação do banco de dados. Neste processo foi aplicado um método de obtenção de chaves para os registros dos dados.

Com a definição de uma estrutura do banco de dados, foram escolhidos três conjuntos de documentos de medição da macrotextura e do coeficiente de atrito nos quais os dados foram extraídos, dois que apresentam resolução adequada e um que apresenta pior qualidade. O *Dataset* apresentado neste trabalho trata-se de uma série de arquivos no formato “.pdf” que contém relatórios de medição de macrotextura e coeficiente de atrito de pistas de pouso e decolagem de alguns aeroportos do Brasil. Tais relatórios foram gerados pelos operadores aeroportuários e apresentados à Agência Nacional de Aviação Civil (ANAC) como um dos critérios de segurança operacional das pistas de pouso e decolagem.

Após a seleção dos documentos teste, foram realizadas as leituras de seus caracteres utilizando o Tesseract OCR. O produto desse processamento são arquivos de texto no formato “.txt” que contém dados extraídos dos relatórios. Contudo, esses dados, geralmente, se apresentam desorganizados e com erros. Desse modo, para que os dados dos arquivos de texto possam ser utilizados de alguma forma, será realizada a limpeza desses dados. Por fim, para validar a capacidade do programa de ler os caracteres de forma correta, foram calculados índices de acurácia das leituras de caracteres para cada página do documento.

4. RESULTADOS PRELIMINARES

A estruturação dos dados escritos nos relatórios de medição do coeficiente de atrito e de macrotextura são de grande importância para as atividades de manutenção e observação dos aeroportos. Para garantir uma melhor trabalhabilidade desses dados pelos usuários, foram analisados os níveis de relevância dos dados contidos nos relatórios apresentados. Com isso, foi elaborado um modelo de planilha no formato “.xls” (Figura 2) que deverá ser um dos *outputs* esperados do algoritmo planejado para este trabalho.

Chave	DATA DE MEDIÇÃO	DATA DA ÚLTIMA REMOÇÃO DE BORRACHA	PISTA DE POUSO E DECOLAGEM	AERONÁVE CRÍTICA DE PROJETO	POUSOS DIÁRIOS	COMPRIMENTO X LARGURA	COMPRIMENTO MEDIDO	PROFUNDIDADE DE MACROTEXTURA				CONDIÇÃO
								1/3	2/3	3/3	MÉDIA	
SBFZ1331MTX200514	20/05/2014	10/05/2014	PISTA DE POUSO E DECOLAGEM 13/31	E	Cabeceira 13 (Predominante): 86	2545 X 45	2500,00	0,87	0,83	0,83	0,85	ADEQUADA
					Cabeceira 31 (Secundária): 1							

Figura 2: Modelo de planilha em “.xls” com os dados de macrotextura

A leitura dos caracteres utilizando o Tesseract OCR, obteve um índice de acurácia favorável na maioria dos casos principalmente entre os relatórios que apresentam resolução adequada (Relatórios 01 e 03), apresentando valores entre 98,01% e 99,45% conforme apresentado na Tabela 2. Contudo, nota-se que em todos os documentos o algoritmo apresenta dificuldades em ler caracteres acentuados, como “ç”, “~” e “’”, além de possuir maior dificuldade de para ler textos grifados em tonalidades de cinza. Além disso, percebeu-se que, em casos de documentos escaneados com baixa resolução, os caracteres de conjunto de chaves (“[” e “]”) são interpretados pelo Tesseract OCR como os caracteres “C” ou “L” ou “J”.

Tabela 1: Resumo dos índices de acurácia obtidos

Página	Relatório de macrotextura	Relatório de macrotextura	Relatório de macrotextura	Relatório de atrito	Relatório de atrito	Relatório de atrito
	01	02	03	01	02	03
Capa	99,45%	92,60%	99,15%	98,14%	96,22%	98,98%
Página 1	98,15%	44,13%	98,01%	98,66%	61,23%	99,41%

5. CONCLUSÕES PRELIMINARES

A organização dos campos de dados é uma etapa importante no desenvolvimento do banco de dados de um sistema de gerência de pavimentos, pois possibilita o entendimento da sua importância aos gestores. Ao organizar esses dados em uma planilha facilita-se o acesso e o uso dos valores de macrotextura e de coeficiente de atrito presente nos relatórios, tornando-os úteis para os operadores aeroportuários e agência reguladora desenvolverem conhecimentos e sabedorias.

A leitura do Tesseract OCR mostrou-se eficiente para textos simples e sem contaminação, porém apresenta erros ao ler documentos com imagens de menor resolução e possui dificuldades de interpretar caracteres acentuados. Para corrigir tais restrições, deve-se prosseguir com um processo de aprendizado de máquina que faça o Tesseract OCR entender esses caracteres. Além disso, o aprendizado de máquina deve fazer o algoritmo interpretar fontes grifadas em escalas de cinza.

Após a sua calibração, espera-se que o Tesseract OCR seja um *software* promissor para realizar a leitura dos documentos. Caso obtenha índices de acurácias superiores, será desenvolvido um algoritmo que conseguirá ler e limpar os dados presentes nos relatórios, organizando-os em planilhas no formato “.xls”.

Agradecimentos

Os autores agradecem o apoio da Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP) que ajudou no desenvolvimento deste trabalho através da concessão de bolsa de pesquisa. Agradecem também à Superintendência de Infraestrutura Aeroportuária (SIA) da Agência Nacional de Aviação Civil pela disponibilização dos dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- Agência Nacional de Aviação Civil, ANAC (2017) Manual de sistema de gerenciamento de pavimentos aeroportuários – SGPA. Superintendência de Infraestrutura Aeroportuária – SIA. Brasília.
- Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; Stein, C. (2009) *Introduction to Algorithms*. MIT Press, 3ª Edição. Cambridge, Massachusetts, EUA.
- Fedeli, R. D.; Polloni, E. G. F.; Peres, F. E. (2010) *Introdução à ciência da computação*. CENGAGE Learning, 2ª Edição. São Paulo, Brasil.
- Fox, P. (2014) *Data Life Cycle: Introduction, Definitions and Considerations*. Notas de Aula. Rensselaer Polytechnic Institute. Nova Iorque, EUA.
- Haas, R.; Hudson, W. R.; Zaniewski, J. P. (1994) *Modern Pavement Management*. Krieger Publishing Company. Malabar, Florida, EUA.
- Hjørland, B. (2018) *Data with big data and database semantics*. Knowledge Organization 45, no. 8: 685-708. ISKO *Encyclopedia of Knowledge Organization*. Disponível em <<http://www.isko.org/cyclo/data>>. Acesso em 5 de junho de 2020.
- Robby, G. A.; Tandra, A.; Susanto, I.; Harefa, J.; Chowanda, A. (2019) *Implementation of Optical Character Recognition using Tesseract with the Javanese Script Target in Android Application*. 4th International Conference on Computer Science and Computational Intelligence. Yogyakarta, Indonesia.
- Zambetti, M.; Pinto, R.; Pezzotta, G. (2019) *Data lifecycle and technology-based opportunities in new Product Service System offering towards a multidimensional framework*. 11th CIRP Conference on Industrial Product-Service Systems. Zhuhai & Hong Kong, China.

Haikel Buganem Busgaib Gonçalves (haikelbusgaib@det.ufc.br)

Francisco Heber Lacerda de Oliveira (heber@det.ufc.br)

Departamento de Engenharia de Transportes, Centro de Tecnologia, Universidade Federal do Ceará
Campus do Pici, BL.703 - Av. Mister Hull, s/n – Fortaleza, CE, Brasil