

**DINCON 2017**CONFERÊNCIA BRASILEIRA DE DINÂMICA, CONTROLE E APLICAÇÕES
30 de outubro a 01 de novembro de 2017 – São José do Rio Preto/SP

Previsão de Gasto Energético de Manobras de Rebocadores Portuários utilizando Máquinas de Vetores Suporte e Florestas Aleatórias

Vladimir da Rocha C. Junior¹Daniel Cruz Cavalieri²Flávio Garcia Pereira³

Programa de Pós-Graduação Stricto Senso em Engenharia de Controle e Automação,
ProPECAut, IFES, Serra, ES

Do ponto de vista da operação de rebocadores portuários, ou gestão do processo, os gastos com combustível podem chegar a mais de 15% dos custos do negócio, o que representa milhões de reais anuais. Assim, sua identificação de forma eficiente representa uma redução de custo significativa, uma vez que este processo pode ocorrer repetidas vezes ao dia. Neste trabalho, a partir de sete características selecionadas da operação dos rebocadores, foram aplicadas técnicas de aprendizado de máquinas para prever qual o possível consumo de combustível nas manobras realizadas. Especificamente, métodos baseados em Máquinas de Vetor Suporte com três *kernels* tradicionais e Florestas Aleatórias kNN foram empregados, tendo seus resultados comparados. Foi demonstrado que as técnicas aplicadas podem alcançar resultados com assertividade de até 77,8% na identificação correta do consumo de combustível.

Palavras-chave. Aciaria, Escória, Processamento digital de imagens, Reconhecimento de padrões

1 Introdução

Rebocadores portuários são embarcações cuja finalidade é prover parte das forças necessárias para atracação e desatracação de navios em portos, evento comumente chamado de manobra pelo meio marítimo e portuário. Desta forma, uma manobra corresponde a um evento de entrada ou saída de um navio em uma bacia ou canal de evolução até sua atracação no porto, como ilustrado na Figura 1. A energia necessária para disponibilizar algumas das

¹ vladimir_cordeiro@terra.com.br

² daniel.cavalieri@ifes.edu.br

³ flavio.garcia@ifes.edu.br

forças envolvidas em uma manobra é proveniente da queima do combustível nos motores das embarcações e sua capacidade de entrega é medida em tração estática, ou *bollard pull*, que é a máxima força contínua que uma embarcação consegue impor em um cabo de reboque sobre condições específicas.

Do ponto de vista da operação de rebocadores portuários, ou gestão do processo, os gastos com energia podem chegar a mais de 15% dos custos do negócio, o que representa milhões de reais anuais, além das crescentes exigências para redução nos níveis de emissão de CO₂ [1]. Logo, entender a dinâmica do sistema é uma condição para a competitividade e a rentabilidade das empresas do ramo. A falta de domínio deste sistema é um problema que dificulta a assertividade de orçamento dos custos, receitas e a melhorias no rendimento do processo.



Figura 1: Manobra de navio em bacia de evolução.

Dentro deste contexto, este trabalho pretende utilizar modelos baseados em algoritmos de aprendizado de máquinas para determinar, a partir de sete características do processo, qual será o possível consumo de combustível em uma manobra realizada por um rebocador. Especificamente, serão utilizadas técnicas de classificação baseadas em Máquinas de Vetores Suportes (do inglês, *Support Vector Machines* - SVM) [7, 11] e combinações de árvores de decisão, conhecidas como Floresta Aleatória (do inglês, *Random Forest* - RF) [2, 10], partindo da premissa de que a análise conjunta se faz necessária dada a capacidade que dados simbólicos têm de influenciar nos desempenhos de algumas técnicas [3].

Não existem muitos trabalhos na literatura sobre este tema utilizando técnicas de aprendizado de máquinas. No entanto, é importante destacar algumas publicações que também tem como objetivo estudar a dinâmica de manobras de atracação e desatracação de navios, da aproximação até a amarração no píer. Em [4] é proposto um modelo matemático que, inserido em um algoritmo de redes neurais artificiais simula o processo e os efeitos das forças envolvidas, a grande contribuição está na possibilidade de procedimentos que otimizem o processo e estabeleçam um padrão, ainda, a possibilidade de sistemas autônomos. Já em [1] são discutidos os índices de emissão de CO₂, com base nas tendências da Organização Marítima Internacional (OMI), visando aspectos de projeto, construção e o tipo de operação, que consequentemente demandam melhora

na eficiência dos equipamentos. O trabalho apresentado em [5] estuda condições operacionais e linhas de base para diversas condições específicas em uma operação de rebocadores portuários estabelecendo parâmetros de monitoramento de condições e operação. Com respeito aos estudos para melhora na eficiência energética deste tipo de operação, poucos ainda são os trabalhos baseados no aprendizado de máquinas por meio Teoria do Aprendizado Estatístico (TAE) ou otimização.

2 Métodos

Antes da aplicação das técnicas propostas neste trabalho é imprescindível estabelecer as variáveis utilizadas no problema tratado. Assim, as variáveis de entrada foram inicialmente mapeadas como parte importante da composição de uma manobra, baseadas no conhecimento de especialistas da gestão do processo, e são apresentadas na Tabela 1.

Como pode ser visto na Tabela 1, a natureza dos dados é variada, indo desde dados numéricos e ordinais, discretos e contínuos, à dados categóricos. Suas distribuições podem ser caracterizadas da seguinte forma:

- Tração estática e DWT (*Dead Weight*) são dados numéricos e contínuos;
- Comandante, Prático, Cais e Posição são dados categóricos convertidos em numéricos pela regra C-to-1;
- Atracação/Desatracação é um dado discreto e binário.

Tabela 1: Atributos de entrada e saída determinados para uma manobra.

Atributos		Tipo
Entrada	Tração Estática	Contínuo
	Comandante	Categórico
	Prático	Categórico
	DWT	Contínuo
	Cais	Categórico
	Desatracação/Atracação	Discreto
	Posição	Categórico
Saída	Consumo de Combustível	Contínuo

Os dados utilizados na análise foram retirados de um documento denominado “Parte Diária”, onde os valores de todos os atributos da Tabela 1 são lançados por cada observação. Inicialmente, os *outliers* foram tratados com base em uma carta de controle estatístico feita sob norma ISO 8258.

Como já dito, as técnicas escolhidas para este trabalho foram as *Random Forest* e as *Support Vector Machines*, ambas implementadas utilizando-se o *software* Matlab®. O algoritmo *Random Forest*, ou floresta aleatória é uma variação das técnicas de junção de classificadores (*ensemble*) formada a partir de classificadores individuais gerados sob observações escolhidas aleatoriamente [2]. Existem diversas variações de técnicas de

formação, como *bagging* e *bootstrap*. Neste estudo será utilizada a técnica de *bagging* em que cada árvore é gerada sobre dados de teste e treinamento aleatórios, o resultado mais votado pelo grupo de classificadores será o escolhido [6]. Entre as vantagens desta técnica podemos destacar o fato de que estes algoritmos são imunes a problemas de *overfitting*. Já as máquinas de vetores vêm recebendo crescente atenção da comunidade de aprendizado de máquinas nos últimos anos, seus resultados são muitas vezes superiores aos obtidos por algoritmos mais tradicionais como Redes Neurais Artificiais [3]. Este método de classificação associado às técnicas de *kernel* e *ensemble* tem excelente generalização e capacidade de trabalhar com dados multivariados e várias classes por meio de algoritmos de correção de erro. Ainda, este conceito pode ser estendido a problemas de regressão com algumas reformulações [7]. Para [3], a técnica possui uma boa capacidade de generalização e robustez diante de objetos de grande dimensão. Para [8], atualmente, cinco variações de SVM estão entre os vinte melhores classificadores, junto com as *Random Forest*.

Desta forma, acredita-se que pela natureza dos dados multivariados com grande quantidade de características simbólicas, as duas técnicas utilizadas aqui são adequadas e podem suportar os objetivos deste trabalho.

Posteriormente, as técnicas de classificação foram aplicadas sobre os dados de saída particionados em faixas de consumo de combustível [9], medidas em litros por hora (L/h). Essa abordagem possibilita a aplicação de um sistema de classificação e um melhor estudo das influências das características de entrada para que a dinâmica do sistema possa ser estudada, inicialmente, sobre a visão de um classificador. Assim, as faixas de saída apresentadas na Tabela 2 foram particionadas obedecendo a dois critérios:

1. Devem estar sujeitas a um domínio que reflita os valores de saída mediante as metas estipuladas pela gestão do processo, a faixa central está dentro dos limites estabelecidos na carta de controle;
2. Devem balancear as distribuições das classes de saída obedecendo ao critério estabelecido na carta de controle.

Tabela 2: Faixas de saída para consumo de combustível, em L/h.

Faixas	Consumo de Combustível (L/h)	Nº de Observações
F1	<95	70
F2	=>95 e <=125	79
F3	>125	44

Para este estudo foram analisadas, inicialmente, 193 operações no Porto de Vitória/Espírito Santo, separadas em etapa de treinamento e teste, particionados na razão de 80/20%, ou seja 80% para treinamento e 20% para teste.

Antes da aplicação das técnicas, os dados das observações foram revistos para possíveis novos tratamentos. A retirada de *outliers* já havia sido executada para os dados disponibilizados, principalmente com base na carta de controle estatístico e nos valores de metas dos indicadores históricos do processo, o foco foi direcionado para os valores distantes da média, porém, admissíveis por estes controles. Após análise junto aos especialistas, foi possível concluir que alguns valores seriam tecnicamente impossíveis de ocorrerem, visto que os equipamentos em regime sem carga só admitiam um limite de 64 l/h, ainda que as observações informassem o contrário. Foram encontrados, também, situações em que os

equipamentos foram utilizados para testes, em condições controladas, que poderiam estar abaixo do mínimo permitido, tais ocorrências foram tratadas também. As intervenções acima reduziram o número de observações para 184 eventos, uma redução de 5% nos dados, mas tornaram os dados mais confiáveis.

Diante dos dados retificados e com base nas faixas definidas na Tabela 2, os algoritmos foram treinados utilizando-se a metodologia citada anteriormente. Para os testes utilizou-se uma abordagem de *holdout* repetida 100 vezes, cujo resultado final foi dado pela média dos testes. Os resultados serão apresentados na próxima seção.

3 Resultados

A avaliação do índice de acerto determinando as faixas foi realizada de duas formas distintas: primeiramente, utilizando-se as 3 faixas determinadas na Tabela 2; e, posteriormente, utilizando-se somente 2 faixas, menor do que 95 L/h e maior do que 95 L/h. Este último teste foi realizado com o intuito de melhorar a predição dos classificadores. É importante destacar que, neste caso, foi mantida a relação de 80/20% para treinamento e teste. Ambos os resultados são apresentados na Tabela 3.

Tabela 3: Resultados obtidos com a aplicação dos classificadores para a predição do consumo de combustível.

Classificadores	Taxa de Acerto Médio (%)	
	3 Faixas	2 Faixas
SVM linear	59,0	66,6
SVM gaussiano	56,0	77,8
SVM polinomial (ordem 3)	57,0	72
RF	57,0	70

Analisando os resultados apresentados na Tabela 3 constata-se um índice de acerto de até 59%, para 3 faixas, utilizando-se o modelo SVM com *kernel* linear. Nesta mesma faixa, o algoritmo RF e a SVM com *kernel* polinomial obtiveram um acerto de 57%, sendo o pior resultado apresentado pela SVM com *kernel* gaussiano com 56% de assertividade. Tais valores ainda estão muito abaixo do desejado para uma aplicação prática. No entanto, quando a faixa de consumo é dividida em duas (<95 L/h e >95 L/h) é notável um aumento da assertividade em todos os algoritmos, sendo o melhor resultado de 77,8% apresentado pela SVM com *kernel* gaussiano. Este resultado já permite uma estimativa aceitável para uma possível aplicação prática.

Mesmo com bons resultados encontrados pelos algoritmos é possível que ainda existam *outliers* no processo, reforçando a necessidade de precisão no apontamento dos dados e melhora nos controles de rotina dos rebocadores. Aqui, é importante destacar a presença dos especialistas para a construção dos algoritmos.

Com o intuito de analisar quais as características possuem maior influência para a predição do consumo de combustível nos rebocadores, foi realizado um teste de importância para 500 simulações. O resultado, também se utilizando o Matlab®, é apresentado na Figura 2 e demonstrou uma forte influência da tração estática (TE), do prático (PILOT) e da posição do rebocador (POS) nas manobras, como já era esperado. As outras variáveis da Figura 1 são: comandante (CMT), *Dead Weight* (DWT), cais (PIER) e atracação/desatracção (AT).

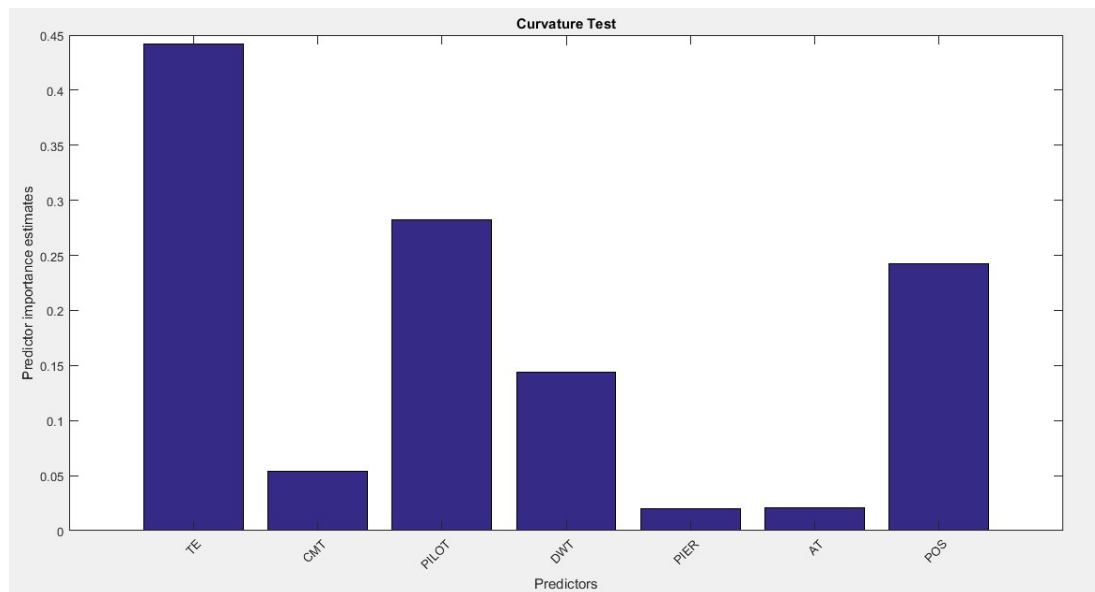


Figura 2: Resposta do teste de importância para as variáveis de entrada do modelo.

4 Conclusões

Do ponto de vista da operação de rebocadores portuários, ou gestão do processo, os gastos com o consumo de combustível nos rebocadores podem chegar a mais de 15% dos custos do negócio, o que representa milhões de reais anuais gastos. Neste sentido, este trabalho propôs a utilização de algoritmos como *Support Vector Machines* e *Random Forest* para determinar, a partir de sete características do processo, qual seriam os possíveis consumos de combustível em 184 manobras realizadas por rebocadores.

Os resultados alcançados pelos métodos aqui propostos demonstraram um índice de acerto para o consumo de combustível de até 77,8%, utilizando-se 2 faixas de consumo. A eliminação de *outliers* foi de fundamental importância para garantir a confiabilidade dos resultados e ajudou, inclusive a melhorar a rotina de operações. Estes resultados mostraram que é possível uma aplicação prática dos métodos utilizados, principalmente utilizando-se o modelo SVM. No entanto, testes com mais manobras e outros algoritmos devem ser realizados para melhor garantir a robustez e aplicabilidade dos métodos aqui propostos.

Agradecimentos

Os autores agradecem ao IFES – Instituto Federal do Espírito Santo, Campus Serra, pelo suporte científico que permitiu desenvolver este trabalho.

Referências

- [1] GASPAR, Henrique M.; ERIKSTAD, Stein Ove. Extending the energy efficiency design index to handle non-transport vessels. In: Proceedings of COMPIT. 2009.

- [2] BREIMAN, Leo. Random forests. *Machine learning*, v. 45, n. 1, p. 5-32, 2001.
- [3] FACELI, Katti, et al. *Inteligência Artificial: Uma abordagem de aprendizado de máquina*. Rio de Janeiro: LTC 2 (2011): 192.
- [4] HASEGAWA, Kazuhiko; FUKUTOMI, T. On harbor maneuvering and neural control system for berthing with tug operation. In: *Proc. of the 3rd International Conference on Maneuvering and Control of Marine Craft*. 1994. p. 197-210.
- [5] TRODDEN, D. G. et al. Fuel usage data analysis for efficient shipping operations. *Ocean Engineering*, v. 110, p. 75-84, 2015.
- [6] LIAW, Andy et al. Classification and regression by Random Forest. *R News*, v. 2, n. 3, p. 18-22, 2002.
- [7] SMOLA, Alex J.; SCHÖLKOPF, Bernhard. A tutorial on support vector regression. *Statistics and computing*, v. 14, n. 3, p. 199-222, 2004.
- [8] FERNÁNDEZ-DELGADO, Manuel et al. Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res*, v. 15, n. 1, p. 3133-3181, 2014.
- [9] GIOLO, Suely Ruiz. *Introdução a Análise de Dados Categóricos*. Curitiba, Universidade Federal do Paraná-Departamento de Estatística, 2006.
- [10] HARTSHORN, Scott. *Machine Learning with Random Forest and Decision Tree*. Kindle Edition, 2016.
- [11] HSU, Chih-Wei et al. *A practical guide to support vector classification*. 2003.