

Estudo comparativo entre modelos de previsão de resistência: leis analíticas *versus* data mining



Mário Coelho*

Aluno Doutoramento
ISISE – Univ. Minho
Guimarães

mcoelho@civil.uminho.pt



José Sena-Cruz

Professor Associado
ISISE – Univ. Minho
Guimarães

jsena@civil.uminho.pt



Luís Neves

Professor Auxiliar
Univ. Nottingham
Nottingham, Reino Unido

luis.neves@nottingham.ac.uk

SUMÁRIO

Neste trabalho apresenta-se um estudo comparativo entre modelos de previsão da resistência de sistemas de polímeros reforçados com fibras (*Fiber Reinforced Polymers – FRP*) inseridos no betão de recobrimento de elementos de betão. Esta técnica é conhecida por *Near-Surface Mounted (NSM)*. São analisados dois dos mais importantes códigos de projeto existentes e ainda modelos obtidos com recurso a algoritmos de *data mining (DM)*. Os resultados obtidos demonstram que o recurso ao DM pode ser uma alternativa viável aos códigos existentes na previsão da resistência da ligação de sistemas FRP NSM.

Palavras-chave: FRP; NSM; Aderência; Analítico; Data Mining

1. INTRODUÇÃO

No contexto do reforço de estruturas de betão, a utilização de polímeros reforçados com fibras (*FRP*) tem surgido nas últimas décadas como uma boa alternativa aos materiais tradicionais. A sua elevada rigidez e resistência à tração, baixo peso, procedimentos de aplicação simples, elevada durabilidade e ainda a disponibilidade praticamente ilimitada em termos de geometrias e tamanhos, são algumas das principais vantagens dos FRP [1].

Uma das técnicas mais promissoras para aplicar FRP em estruturas de betão é a designada técnica de reforço NSM (*Near-Surface Mounted*), que consiste na instalação de laminados ou varões de FRP em entalhes abertos no betão de recobrimento [2,3].

Duas das mais importantes propostas para a previsão da resistência de sistemas FRP NSM encontram-se presentes nas normas 440.2R-08 do *American Concrete Institute* (ACI) [4] e HB 305-2008 da *Standards Australia* (SA) [5]. Ambas se baseiam em análises de regressão de resultados experimentais de ensaios de arranque presentes na literatura. No entanto, os parâmetros de entrada utilizados por cada proposta são diferentes, o que reflete a incerteza que ainda existe sobre os parâmetros que mais influência têm e/ou podem descrever melhor os fenómenos complexos associados à técnica FRP NSM. Por outro lado, o simples facto de utilizarem uma única expressão, com um número limitado de parâmetros de entrada, para descrever a resistência da ligação FRP NSM, introduz por si só um certo nível de erro. De modo a tentar resolver problemas idênticos a este, a utilização de algoritmos de *Data Mining* (DM) tem revelado resultados promissores noutras áreas da Engenharia Civil [6,7]. No entanto, os autores do presente trabalho não encontraram na literatura referências a aplicação de algoritmos de DM no âmbito do reforço de elementos de betão com FRP NSM.

DM é uma área relativamente nova da ciência da computação que combina diferentes técnicas como a estatística, a inteligência artificial e a gestão de dados, entre outras, para proporcionar um conhecimento mais profundo sobre os dados presentes numa determinada base de dados. Assim, o DM está a emergir como uma classe de técnicas numéricas que vão além da estatística e que se preocupa com descobrir, simplificar e resumir, de forma automática, padrões e/ou correlações dentro do conjunto de dados em análise [8].

Neste trabalho, são analisadas, para além dos dois códigos de projeto referidos anteriormente, modelos de previsão gerados por dois algoritmos de DM, nomeadamente, as redes neuronais artificiais (*Artificial Neural Networks* – ANN) e as máquinas de vetores de suporte (*Support Vector Machines* – SVM) [9].

2. MODELOS ANALÍTICOS DE PREVISÃO E ALGORITMOS DE DM

Nos próximos parágrafos apresentam-se de forma sucinta as propostas ACI e SA, os algoritmos de DM utilizados e ainda a base de dados que permitiu comparar as propostas analíticas com as baseadas nos algoritmos de DM.

2.1 Leis analíticas existentes

As duas formulações existentes para a análise de sistemas FRP NSM abordadas neste trabalho, ACI e SA, baseiam-se na mesma filosofia: se o comprimento da ligação for suficiente (comprimento de desenvolvimento – L_d), toda a resistência da ligação será mobilizada, caso contrário, esta será reduzida linearmente de acordo com o comprimento da ligação real (L_b).

No Quadro 1 apresentam-se os principais aspetos das formulações ACI e SA. A definição dos parâmetros geométricos de interesse é apresentada na Figura 1. O significado de cada parâmetro é o seguinte:

- A_f, p_f – Área e perímetro da secção transversal do FRP;
- b_f, b_g – Largura do FRP e do entalhe, respetivamente;
- d_f, d_g – Profundidade do FRP e do entalhe, respetivamente;
- E_f – Módulo de elasticidade do FRP;
- f_{fd} – Resistência à tração de cálculo do FRP;
- f_c – Resistência à compressão de cálculo do betão;
- τ_{med}, τ_{max} – Tensões de aderência média e máxima, respetivamente;
- L_{per} – Perímetro da superfície de rutura (afastado 1 mm do perímetro do entalhe).

Quadro 1. Leis analíticas para a avaliação da resistência de sistemas FRP NSM.

Lei analítica	Comprimento de desenvolvimento [L_d]	Força de arranque máxima [F_{tmax}]	Comentários
ACI	$\frac{A_f f_{fd}}{p_f \tau_{med}}$	$A_f f_{fd}$	$\tau_{med} = 6.9 \text{ MPa}$
SA	$2 \sqrt{\frac{\tau_{max} L_{per}}{\delta_{max} (EA)_f}}$	$\sqrt{\tau_{max} \delta_{max} L_{per} (EA)_f} \leq A_f f_{fd}$	$\tau_{max} = (0.8 + 0.078 \varphi_{per}) f_c^{0.6}$ $\tau_{max} \delta_{max} = 0.73 \varphi_{per}^{0.5} f_c^{0.67}$ $\varphi_{per} = d_{per} / b_{per}$

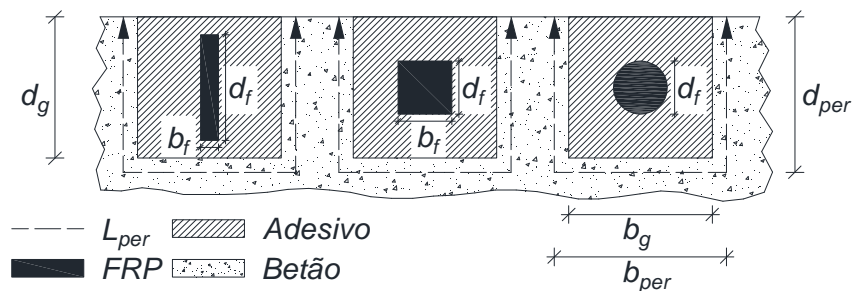


Figura 1. Parâmetros geométricos utilizados nas leis analíticas.

Conforme se pode constatar, a formulação do ACI apresenta um forma bastante mais simples do que a da SA, materializada pelo número reduzido de parâmetros requeridos e tipo de equações bastante simples. Assim, no caso do ACI são necessários apenas 4 parâmetros (L_b, p_f, A_f, f_{fd}) enquanto no caso da SA são necessários 7 parâmetros ($L_b, A_f, f_{fd}, d_g, b_g, E_f, f_c$).

2.2 Data mining

O *data mining* (DM) é uma parte do processo de descoberta de conhecimento útil a partir de bases de dados. Consiste na aplicação de algoritmos específicos para extração de modelos a partir dos dados existentes, mesmo quando estes aparentemente não estão relacionados.

Nesse contexto, os algoritmos de DM usam dados reais a partir de uma base de dados com exemplos do fenómeno em estudo. Através de um processo de aprendizagem interativo, os algoritmos são capazes de prever o comportamento do fenómeno em causa quando novos dados lhe são fornecidos.

Os algoritmos testados e analisados neste trabalho foram as ANN e SVM. Foi adotado o programa R [10], um programa estatístico gratuito e de código aberto. Foi ainda utilizada uma biblioteca do R, especificamente apta para aplicação dos algoritmos ANN e SVM, designada RMiner [9].

Os algoritmos ANN e SVM são descritos na literatura como *black box* ou algoritmos de “caixa preta”, uma vez que não permitem obter uma equação que relaciona as variáveis de entrada e de saída, tal como acontece com as formulações analíticas do ACI e SA. No entanto, os algoritmos de DM são capazes de aprender o fenómeno usando um conjunto de dados de treino e conseguem identificar as relações intrínsecas e complexas (se existirem) entre as várias variáveis, fornecendo assim um modelo de previsão. Adicionalmente, a biblioteca RMiner permite que estes algoritmos forneçam a importância relativa que cada um dos parâmetros de entrada tem nos resultados obtidos. Esta propriedade permite uma compreensão mais profunda das relações entre os diferentes parâmetros, por um lado, mas também permite analisar se essas relações são significativas e/ou fazem sentido fisicamente.

Neste trabalho, a base de dados de ensaios de arranque direto utilizada foi dividida em quatro partes iguais. Depois, foram definidos dois subgrupos com $\frac{3}{4}$ e $\frac{1}{4}$ dos dados. O primeiro subgrupo foi utilizado como conjunto de treino, no qual o modelo de previsão é obtido, enquanto o segundo foi utilizado para testar o modelo obtido. Este processo repete-se para as várias combinações de $\frac{3}{4}$ treino + $\frac{1}{4}$ teste das quatro partes definidas inicialmente. De cada vez que isto é feito, um modelo de previsão é gerado, sendo o modelo final obtido aquele que apresente menores erros de previsão no conjunto treino + teste.

De modo a facilitar a comparação entre os modelos de previsão obtidos pelos algoritmos de DM e as leis analíticas, foram consideradas duas análises distintas. A primeira adotando os mesmos parâmetros de entrada que a formulação ACI. A segunda utilizando os mesmo parâmetros de entrada que a formulação SA.

2.3 Base de dados

A base de dados com os ensaios de arranque direto utilizada neste trabalho contém 363 registos, i.e. informação relativa a 363 ensaios de arranque direto compilados na literatura [3]. Destes, 308, 28 e 27 utilizavam FRP com fibras de carbono, vidro e basalto,

respetivamente. Em termos de secção transversal do FRP, a base continha 175, 35 e 153 registos utilizando FRP com secção retangular, quadrada e redonda, respetivamente. Em todos os registos o adesivo aplicado era do tipo epóxi.

No Quadro 2 apresenta-se um resumo do intervalo de valores dos parâmetros utilizados pelas formulações ACI e SA. Uma vez que esses também foram adotados nos modelos de DM, ficam assim estabelecidos os intervalos de valores de cada parâmetro de modo a que a utilização dos modelos aqui apresentados possa ser a mais adequada. Todas as análises a apresentar no que se segue deste trabalho tiveram como suporte esta base de dados.

Conforme definido na secção 2.1, os parâmetros de entrada diferem entre as formulações ACI e SA. No entanto, por vezes os autores dos trabalhos experimentais não disponibilizam todos os parâmetros necessários. Assim, no que se segue deste trabalho, as análises com os parâmetros de entrada do ACI foram realizadas usando os 363 registos existentes na base de dados. As análises com os parâmetros de entrada da SA foram realizadas apenas com 286 registos, pois os restantes não apresentavam todos os parâmetros necessários.

Quadro 2. Resumo dos principais parâmetros existentes na base de dados e utilizados neste trabalho.

Parâmetro	Número de registos	Intervalo de valores	Parâmetro	Número de registos	Intervalo de valores
A_f [mm ²]	363	[12-201.06]	f_c [MPa]	309	[18.4-65.7]
b_g [mm]	340	[3-50]	f_f [MPa]	363	[512-3100]
d_g [mm]	359	[5-60]	L_b [mm]	363	[30-510]
E_f [GPa]	361	[37.17-273]	ρ_f [mm]	363	[18.85-84.8]

3. RESULTADOS

Os próximos parágrafos apresentam os principais resultados obtidos na análise efetuada, tanto com as formulações ACI e SA, como com os algoritmos de DM. Em toda a análise, foram considerados os valores médios das propriedades mecânicas dos materiais, sendo que nenhum fator de segurança e/ou redução de força foi adotado. Assim, admite-se que a comparação entre os resultados experimentais obtidos em laboratório e as previsões de cada formulação foi a mais adequada. Esta condição implica que, comparativamente aos parâmetros apresentados na secção 2.1, apenas mudem a resistência à tração e compressão do FRP e do betão, respetivamente. A primeira surge em ambas formulações ACI e SA com o seu valor de projeto (f_{td}) mas no que se segue foi considerada com o seu valor último obtido, sempre que possível, em ensaios de tração uniaxial feitos pelos autores dos trabalhos ou nas fichas técnicas dos respetivos FRP (f_f). A segunda surge na formulação SA com o seu valor de projeto (f_c) mas foi considerada através do seu valor médio, obtido pelos autores dos respetivos trabalhos experimentais em ensaios de compressão uniaxial (f_{cm}).

A Figura 2 apresenta uma comparação entre a força de arranque máxima obtida experimentalmente ($F_{fmax,Exp}$) e aquela obtida numericamente ($F_{fmax,Num}$) com formulações que utilizam os parâmetros de entrada da formulação do ACI. Isto inclui, para além da própria formulação do ACI, os modelos com ANN e SVM.

De modo a simplificar a análise dos resultados, foram adicionados a todos estes gráficos três linhas a traço interrompido. A do meio corresponde à linha de 45°, que determina os cenários em que a previsão numérica coincide com o resultado experimental. As restantes duas linhas delimitam as fronteiras de variação de 20%. Assim, as ocorrências que se situam acima da linha de 45° indicam que a formulação subestima o resultado experimental, enquanto as ocorrências abaixo sobrestimam o resultado experimental. O valor de 20% foi considerado como um valor razoável para ter em conta todos os desvios que pudessem existir em relação aos experimentais disponíveis.

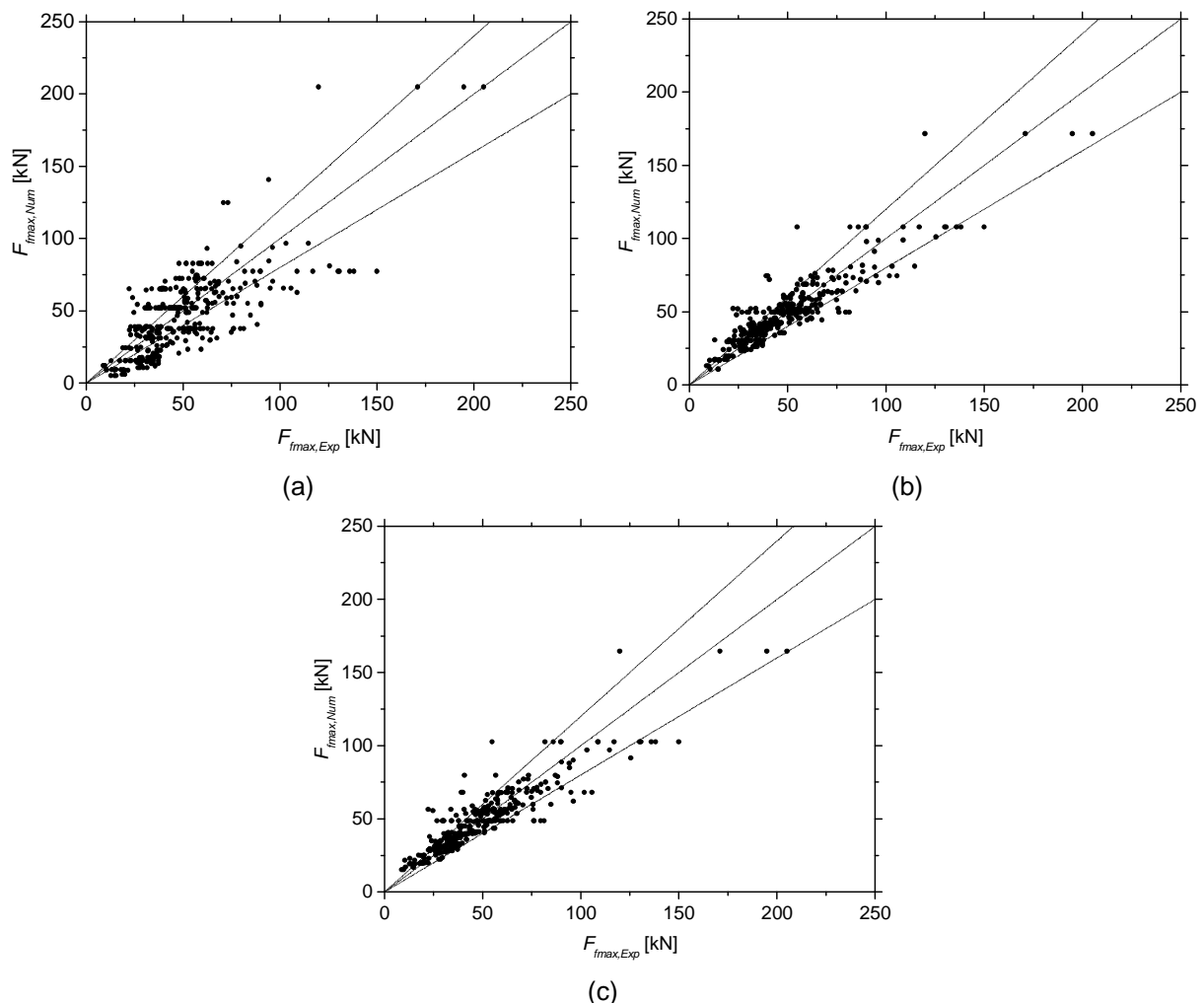


Figura 2. Comparação da máxima força de arranque em termos experimental *versus* previsão numérica usando os parâmetros de entrada da formulação ACI (363 registos) e: (a) a formulação prevista no ACI; (b) o modelo ANN; (c) o modelo SVM.

Na Figura 3 apresenta-se o mesmo tipo de resultados, mas para as análises que utilizaram os parâmetros de entrada da formulação da SA. Igualmente, para além da formulação SA, apresentam-se os resultados de modelos com ANN e SVM.

Conforme é possível constatar (para ambas as figuras) a capacidade preditiva dos modelos de DM é melhor comparativamente às formulações analíticas previstas nos códigos. De facto, a dispersão de resultados verificada com a utilização das formulações analíticas é substituída por uma maior concentração de resultados na região delimitada pelas linhas de variação $\pm 20\%$ quando se utilizam os modelos de DM. Não obstante a significativa melhoria nas previsões, os modelos baseados em algoritmos de DM continuam a apresentar algumas previsões fora da região de erro até 20%.

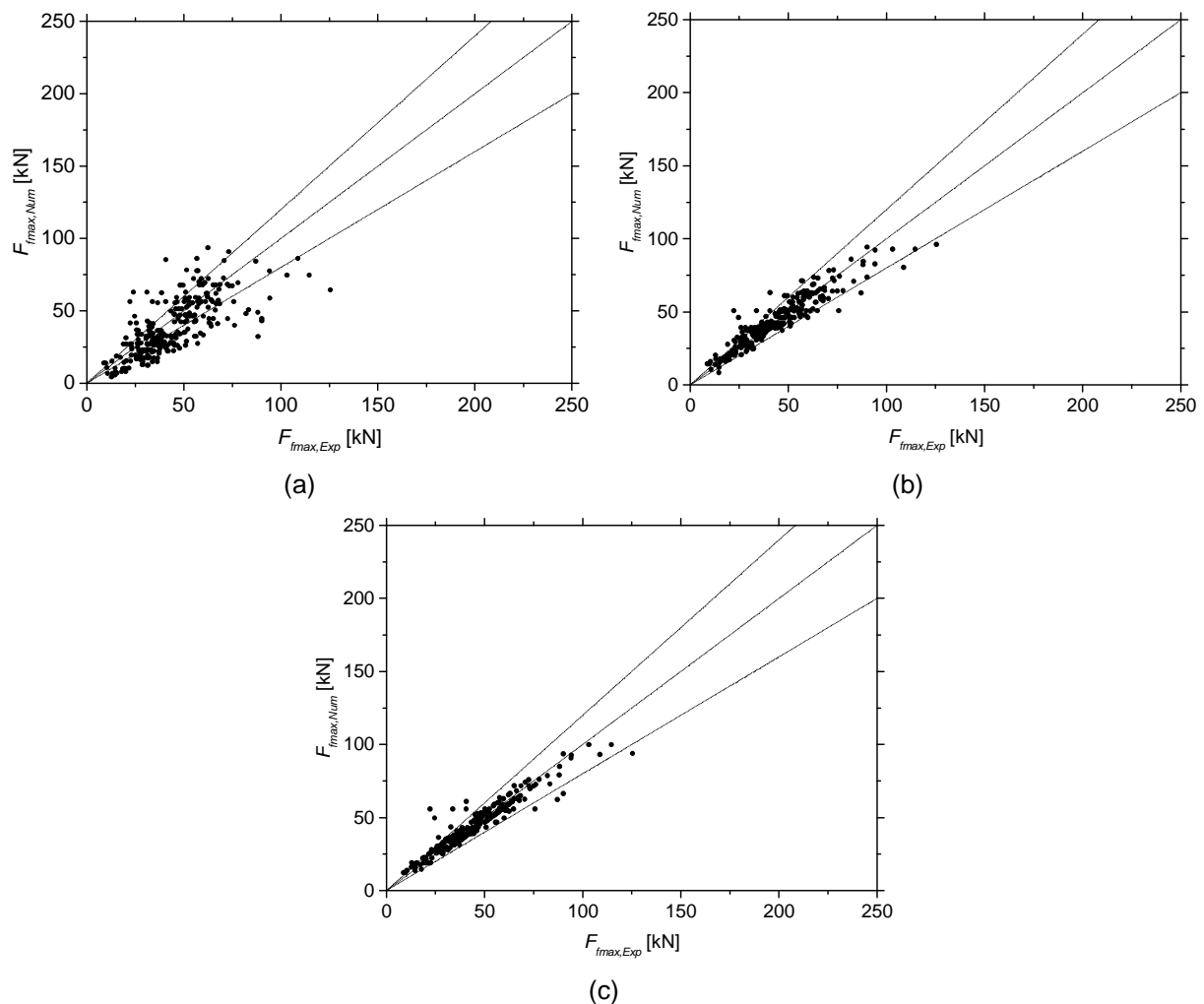


Figura 3. Comparação da máxima força de arranque em termos experimental *versus* previsão numérica usando os parâmetros de entrada da formulação SA (286 registos) e: (a) a formulação prevista na SA; (b) o modelo ANN; (c) o modelo SVM.

Outro aspeto que merece destaque é o facto de, em ambas as formulações analíticas, os resultados fora da região de erro até 20% surgirem maioritariamente na zona inferior do gráfico. Isto significa que, apesar de ainda apresentarem algumas imprecisões de previsão, ambas as formulações ACI e SA tendem a ser conservativas nas suas previsões.

Para além da informação que se pôde obter nas figuras anteriores, a avaliação da capacidade preditiva das várias formulações em análise foi realizada através do cálculo de algumas medidas de erro. O rácio δ_i entre a previsão numérica e valor experimental da força de arranque máxima, obtidas para cada ensaio de arranque existente na base de dados, foi selecionado como parâmetro de comparação (1). Para este rácio o valor mínimo (MIN), médio (MED), máximo (MAX), desvio padrão (DP) e coeficiente de variação (CV) foram calculados.

Além dessas medidas de erro tradicionais, foi ainda estimado o erro médio absoluto (MAE) e o erro quadrático médio (RMSE) para cada análise. A equação (2) apresenta a forma de estimativa do erro, enquanto as expressões (3) e (4) apresentam a forma como MAE e RMSE foram obtidas neste trabalho.

Por último, foi também calculado o coeficiente de correlação quadrática (R^2) para todas as análises com base nos valores das previsões e experimentais da máxima força de arranque.

$$\delta_i = (F_{f \max, Num} / F_{f \max, Exp})_i \quad (1)$$

$$e_i = (F_{f \max, Num})_i - (F_{f \max, Exp})_i \quad (2)$$

$$MAE = \sum_{i=1}^N |e_i| / N \quad (3)$$

$$RMSE = \sqrt{\sum_{i=1}^N e_i^2 / N} \quad (4)$$

No Quadro 3 apresentam-se os valores das referidas medidas de erro, obtidas para as três análises que utilizavam os parâmetros de entrada do ACI. Neste quadro, a primeira análise corresponde à formulação do ACI conforme apresentada na secção 2.1, sendo as restantes duas correspondentes aos resultados dos modelos de DM utilizando os algoritmos ANN e SVM. Assim, os resultados apresentados neste quadro para ACI, ANN e SVM dizem respeito à Figura 2a, b e c, respetivamente.

No Quadro 4 apresentam-se as mesmas medidas de erro para as análises que utilizavam os parâmetros de entrada da SA. Analogamente, os resultados apresentados neste quadro para SA, ANN e SVM dizem respeito à Figura 3a, b e c, respetivamente.

Da análise destes dois quadros foi possível, uma vez mais, verificar a maior capacidade preditiva dos modelos de DM desenvolvidos. Este tipo de informação numérica permite, contrariamente à informação gráfica apresentada nas figuras anteriores, estabelecer um

ranking dos modelos de previsão. Assim, por ordem crescente de precisão das previsões, surge a formulação original do ACI, a formulação original da SA, os modelos ANN e SVM usando os parâmetros de entrada do ACI e, por último, os modelos ANN e SVM usando os parâmetros de entrada da SA.

Este *ranking* dos modelos de previsão afigura-se razoável uma vez que, tal como anteriormente se havia referido, a formulação do ACI é mais simples que a da SA. Essa simplicidade resulta por isso numa maior dificuldade de prever um fenómeno tão complexo. Por outro lado, sabendo-se que o algoritmo SVM é, comparativamente com o algoritmo ANN, mais robusto, seria expectável que, para o mesmo tipo de parâmetros de entrada, as suas previsões fossem mais precisas. Tal facto é também corroborado de forma explícita nas medidas de erro calculadas para as análises realizadas.

Quadro 3. Medidas de erro das análises utilizando os parâmetros de entrada da formulação ACI (363 registos).

Análise	MIN [-]	MED [-]	MAX [-]	DP [-]	CV [%]	MAE [kN]	RMSE [kN]	R ² [-]
ACI	0.30	0.94	2.94	0.39	41.79	14.85	19.34	0.58
ANN	0.61	1.05	2.37	0.25	24.20	7.94	11.45	0.82
SVM	0.60	1.05	2.54	0.24	23.12	7.75	11.27	0.82

Quadro 4. Medidas de erro das análises utilizando os parâmetros de entrada da formulação SA (286 registos).

Análise	MIN [-]	MED [-]	MAX [-]	DP [-]	CV [%]	MAE [kN]	RMSE [kN]	R ² [-]
SA	0.33	0.89	2.62	0.33	37.72	11.56	15.16	0.53
ANN	0.58	1.03	2.29	0.18	17.43	4.79	6.87	0.87
SVM	0.72	1.03	2.51	0.16	15.58	3.87	5.77	0.91

Conforme referido anteriormente, os algoritmos de DM não só permitem obter a previsão da máxima força de arranque, como também permitem saber a influência de cada parâmetro de entrada na previsão da mesma. Esta foi obtida através de uma análise de sensibilidade que, recorrendo à biblioteca RMiner, foi realizada automaticamente para cada algoritmo. O resultado desta análise de sensibilidade é o peso relativo de cada parâmetro de entrada.

Na Figura 4 apresentam-se os resultados obtidos na análise de sensibilidade realizada sobre os modelos de DM obtidos quando se utilizam os mesmos parâmetros de entrada que as formulações ACI e SA. Esta figura permite verificar que o comprimento da ligação é o parâmetro mais importante, seguido da área do FRP. Numa análise mais genérica pode-se mesmo dizer que os parâmetros de cariz geométrico se revelam mais influentes do que os

parâmetros de natureza mecânica, i.e. as propriedades dos materiais. A explicação mais razoável para esta diferença poderá estar relacionada com os modos de rotura obtidos experimentalmente. De fato, é expectável que a resistência à tração do FRP tenha uma forte influência na máxima força de arranque, quando o modo de rotura é precisamente a rotura do FRP. De igual modo, é expectável que a resistência à compressão do betão seja preponderante quando a rotura é coesiva ao longo da camada de betão próxima do sistema adesivo/FRP. No entanto, para os restantes modos de rotura possíveis, a influência destas duas resistências não deverá ser tão importante. Assim, tendo na mesma base de dados, ensaios de arranque com modos de rotura distintos, os resultados desta análise de sensibilidade devem ser analisados com especial cuidado.

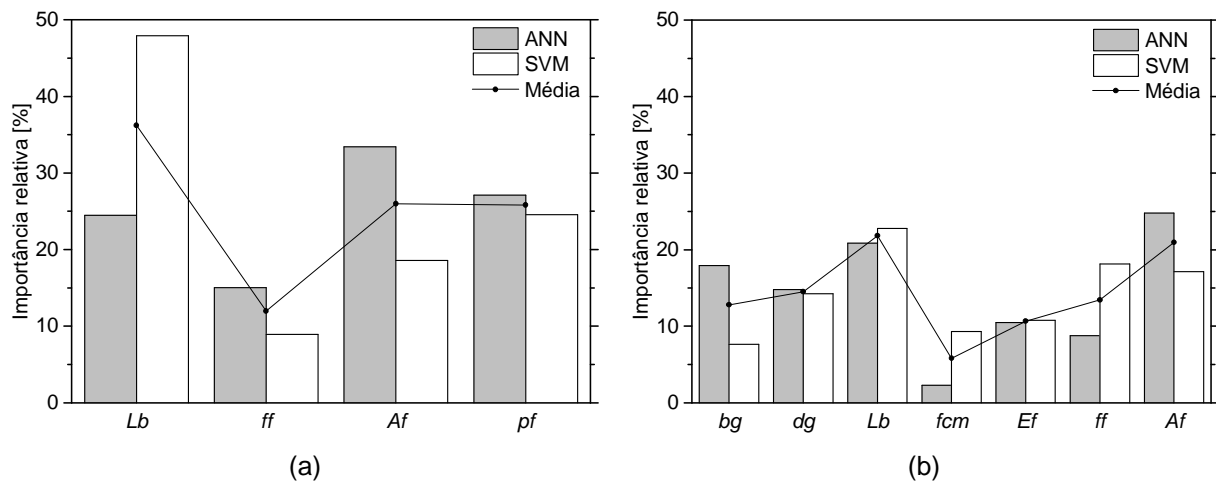


Figura 4. Importâncias relativas obtidas pelos dois algoritmos de DM usando os parâmetros de entrada da formulação: (a) ACI; (b) SA.

4. CONCLUSÕES

Neste trabalho apresentou-se uma análise comparativa entre os modelos existentes de previsão da força de arranque máxima em sistemas de FRP NSM e modelos alternativos baseados em algoritmos de DM.

Os resultados obtidos revelam que são necessárias melhorias nos modelos existentes, uma vez que estes apresentam uma grande dispersão de previsões. Os resultados também revelam que os algoritmos de DM podem ser efetivamente uma boa alternativa aos modelos existentes.

Este trabalho permitiu também confirmar que a formulação mais simplista do ACI conduz a previsões piores do que a formulação mais complexa da SA. Não obstante, convém ter em conta que o número de registos da base de dados utilizado para as análises com os parâmetros do ACI era maior que os utilizados para análises com os parâmetros da SA.

Comparando diretamente os dois algoritmos de DM testados, ficou igualmente verificada a melhor capacidade de previsão do algoritmo SVM.

AGRADECIMENTOS

Este trabalho foi apoiado por fundos FEDER através do Programa Operacional Fatores de Competitividade – COMPETE e de fundos nacionais através da FCT – Fundação para a Ciência e Tecnologia no âmbito do projeto CutInDur PTDC/ECM/112396/2009. O primeiro autor gostaria também de agradecer a bolsa de doutoramento número SFRH/BD/87443/2012 concedida pela FCT.

REFERÊNCIAS

- [1] Bakis et al. – “Fiber reinforced polymer composites for construction-state-of-the-art review”, *Journal of Composites for Construction*, 2002, 6(2):73–87.
- [2] De Lorenzis, L.; Teng, J.G. – “Near-surface mounted FRP reinforcement: an emerging technique for strengthening structures”, *Composites Part B*, 2007, 38(2):119–43.
- [3] Coelho et al. – “A review on the bond behavior of FRP NSM systems in concrete”, *Construction and Building Materials*, 2014, (em revisão).
- [4] ACI 440F – “Guide for the design and construction of externally bonded FRP systems for strengthening concrete structures”, *Report 440.2R-08*, American Concrete Institute, USA, 2008, 80 p.
- [5] SA – “Design handbook for RC structures retrofitted with FRP and metal plates: beams and slabs”, *HB 305-2008*, Standards Australia, Australia, 2008, 76 p.
- [6] Tinoco, J.; Correia, A.; Cortez, P. – “Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time”, *Construction and Building Materials*, 2011, 25:1257–1262
- [7] Martins, F.; Miranda, T. – “Estimation of the Rock Deformation Modulus and RMR Based on Data Mining Techniques”, *Geotechnical and Geological Engineering*, 2012, 30(4):787-801.
- [8] Fayyad, et al. – “From Data Mining to Knowledge Discovery: an overview”, In Fayyad (editor), *Advances in Knowledge Discovery and Data Mining*. AAAI Press / The MIT Press, Cambridge, 1996, p. 471-493.
- [9] Cortez, P. – “Data mining with neural networks and support vector machines using the R/rminer tool”, In: Perner P (ed) *Advances in data mining. Applications and theoretical aspects*. Proceedings of 10th industrial conference on data mining, Lecture Notes in Computer Science, Springer, Berlin, 2010, p. 572–583.
- [10] R Development Core Team, “R: A language and environment for statistical computing”, *R Foundation for Statistical Computing*, Vienna, Austria, (2010). website: <http://www.R-project.org>.