

AVALIAÇÃO DA PERFORMANCE DAS REDES NEURONAIS ARTIFICIAIS E DOS MODELOS DE ESCOLHA DISCRETA NA AQUISIÇÃO DE PRODUTOS COM ENVOLVIMENTO FRACO

Paulo Alexandre Botelho Rodrigues Pires

RESUMO

Neste trabalho comparamos a performance das redes neuronais artificiais com a performance dos modelos de escolha discreta na previsão da aquisição de produtos com envolvimento fraco. Dentro da classe dos modelos de escolha discreta restringimos a comparação ao modelo Logit Multinomial, o modelo base e também o mais simples. Para além desse modelo, considerámos também o Mixed Logit, que é o mais completo e flexível modelo de escolha discreta, conseguindo aproximar as probabilidades de qualquer outro modelo de escolha discreta baseado na maximização da utilidade. A performance dos modelos foi aferida numa base de dados contendo o registo de aquisições de produtos em supermercados. Os resultados obtidos nas execuções das simulações mostraram que não existe uma clara supremacia de um ou outro tipo de modelos. No entanto, os modelos de escolha discreta foram sempre mais robustos e menos exigentes em recursos computacionais

PALAVRAS CHAVE: Redes Neuronais Artificiais, Multinomial Logit, Mixed Logit.

ABSTRACT

In this work, we compared the performance of the artificial neuronal networks with the performance of discrete choice models in forecasting acquisition of products with low involvement. Among the class of discrete choice models, we have restricted the comparison to the Multinomial Logit model, which is the base model and the simpler one. We also evaluated the Mixed Logit model, which is the more complete and flexible discrete choice model, and it can approximate the probabilities of any discrete choice model based on the random utility maximization. The performance of the models was evaluated in a database containing records of products purchases in supermarkets. The results obtained from the simulations executions revealed that it does not exist an evident supremacy of one type of models over the others. However, discrete choice models were always more robust and less demanding in computational resources.

KEY WORDS: Artificial Neural Networks, Multinomial Logit, Mixed Logit.

1. INTRODUÇÃO

Na sociedade actual a disponibilidade e o acesso à informação sobre o comportamento dos consumidores é uma realidade irreversível. Os avanços tecnológicos continuarão a sustentar e a aumentar esse fluxo de informação que resultará numa hipotética situação ideal para os gestores de marketing. Por si só, a abundância e a acessibilidade da informação não garantem a interpretação mais útil da mesma. Para esse efeito, o gestor de

marketing tem de apurar modelos, cuja análise lhe permitirá tomar as decisões mais adequadas. Esta investigação assume como motivação principal o estudo de modelos que permitam incluir a informação proveniente do comportamento do consumidor, das suas características, das suas escolhas e, posteriormente, aferir comportamentos futuros. O seu âmbito restringiu-se a duas áreas disciplinares: modelos de escolha discreta (MED) e redes neuronais artificiais (RNA).

Comprovada a proliferação de paradigmas, a abrangência do estudo teve que ser limitada a um número mais reduzido. Assim, nos MED o estudo centrou-se no modelo Logit Multinomial, considerado o modelo base e o mais simples, e no modelo Mixed Logit, o último avanço científico e o representante de uma classe de modelos flexíveis e mais poderosos. Pretendeu-se, pois, avaliar e comparar o modelo mais simples com o modelo mais poderoso, confirmando se os recursos e a complexidade inerentes ao segundo modelo justificam a sua aplicação em detrimento do primeiro. Nas RNA o estudo restringiu-se às arquitecturas unidireccionais, sintetizando as variantes do algoritmo retropropagação (BP), os algoritmos construtivos e de eliminação de parâmetros, as Radial Basis Function (RBF) e a integração dos algoritmos genéticos (AG) e das RNA.

Depois de introduzido o tema da investigação, elucidando a sua relevância para a área do marketing, indicam-se a seguir os seus objectivos. O objectivo principal incidiu na avaliação da performance das RNA e dos MED, tentando identificar qual dos dois conjuntos de modelos é melhor, isto é, se as RNA têm uma performance superior aos MED na previsão do comportamento do consumidor ou o contrário. Associado a este primeiro objectivo procurou, similarmemente, identificar-se qual dos dois conjuntos de modelos é mais robusto.

Formularam-se ainda outros objectivos pertinentes, inseridos numa perspectiva mais metodológica, inerentes à calibração das RNA. Refira-se especificamente os seguintes objectivos: inicialização dos valores das conexões; pré-processamento das variáveis explicativas; escolha do tipo de aprendizagem; escolha da função erro; utilização de regularização; aplicação da paragem antecipada da aprendizagem. Na generalidade dos estudos efectuados e aqui revistos estes aspectos são, quase sempre, ignorados ou apresentam-se pouco esclarecedores. De facto, não foi encontrado nenhum estudo que avaliasse todos os aspectos enumerados atrás, sendo de acentuar que nas fontes de informação específicas às RNA os mesmos são considerados fundamentais.

A problemática da inicialização dos valores das conexões centra-se na escolha de uma alternativa entre duas possíveis: optar por um procedimento de inicialização simples, em que os valores são inicializados aleatoriamente com valores pequenos, ou optar por procedimentos de inicialização mais complexos mas superiores. O tipo de aprendizagem concentra-se também numa dicotomia, aprendizagem incremental e batch, em que cada vertente suscita vantagens sobre a oposta. A mesma situação é reconhecida para as funções erro, em que a escolha conflui para duas alternativas: função erro média da soma do erro quadrado ou função erro entrópica. A revisão das fontes de informação destaca a maior utilização da função erro média da soma do erro quadrado, mas, para problemas de classificação, a função erro mais adequada é a função erro entrópica.

Como são modelos não paramétricos, as RNA seguem geralmente procedimentos heurísticos na sua calibração. A ocorrência de sobre-aprendizagem é um dos problemas principais a evitar, pois o incremento contínuo do número de parâmetros proporciona uma diminuição do erro, ou seja, um ajuste mais perfeito aos dados da aprendizagem, mas esse incremento contínuo do número de parâmetros tem como consequência uma diminuição da generalização da rede neuronal artificial. Esta dificuldade é atenuada com a inclusão de técnicas como a utilização de regularização ou a paragem antecipada da aprendizagem.

Seguindo a mesma orientação, comprova-se identicamente uma predominância quase total das variantes do algoritmo BP sobre todos os outros paradigmas das RNA. Considerando que nos problemas de classificação nas fontes de informação das RNA certos paradigmas demonstram uma performance superior às variantes do algoritmo BP é importante avaliar a performance dos mesmos na previsão do comportamento do consumidor. Também nas variantes do algoritmo BP se confirma a predominância de alguns algoritmos com destaque para o algoritmo BP e outros como o algoritmo com coeficientes de aprendizagem variáveis e o algoritmo com momentum. Da mesma forma é igualmente importante avaliar a performance dos algoritmos citados nas fontes

de informação como mais rápidos, de que são exemplos os algoritmos baseados em otimização numérica de segunda ordem e os algoritmos baseados em heurísticas. No caso específico dos algoritmos baseados em otimização numérica de segunda ordem é indispensável confirmar a sua tendência para convergirem para mínimos locais fracos, referida inúmeras vezes nas fontes de informação.

A avaliação e comparação dos dois conjuntos de modelos foram aferidas empiricamente, tendo sido executadas várias simulações, as quais obtiveram medidas de desempenho, previamente definidas, que permitiram identificar os modelos melhores. Os dois conjuntos de modelos foram avaliados e comparados utilizando-se uma base de dados de aquisições de produtos com um envolvimento fraco, nomeadamente a aquisição de detergentes líquidos em supermercados. Essa base de dados foi retirada de um estudo de mercado elaborado especificamente para estudar o comportamento dos consumidores nas aquisições de produtos com envolvimento fraco em supermercados. Tendo o estudo sido desenhado por especialistas, contém toda a informação necessária a esta investigação, com incidência particular na identificação das variáveis explicativas e nas alternativas escolhidas. É de destacar que a informação se refere a aquisições autênticas, garantindo assim uma avaliação verdadeira da capacidade previsional dos modelos.

2. REVISÃO DAS FONTES DE INFORMAÇÃO

Uma das áreas de marketing em que o estudo da comparação das RNA e dos MED tem sido, continuamente, investigado é a escolha da marca. Os parágrafos seguintes contêm uma síntese das fontes de informação mais recentes.

O modelo Logit Multinomial e as RNA foram comparados no contexto de envolvimento fraco na compra de produtos em supermercados, Agrawal e Schorling (1996). Os modelos foram comparados com dados provenientes de um ponto de venda, evitando a inclusão de variáveis representativas do “servicescape” (destaque, posição de venda, etc.). Os dados dos grupos foram sempre divididos em dois subconjuntos: estimação e previsão para o modelo Logit Multinomial ou aprendizagem e teste para a RNA. As proposições elaboradas pelos autores não deixam de ser relevantes e comprovar e contradizer estudos anteriores: (1) as RNA têm uma capacidade previsional superior ao modelo Logit Multinomial na previsão de quotas de marcas, quando o problema é complexo para o modelo previsional, que é o caso em que o número de marcas de um produto é grande; (2) as RNA são superiores na identificação das variáveis de entrada cujo efeito é estatisticamente negligível nas variáveis de saída para clientes que têm uma fidelização à marca grande; (3) nas situações de troca de marca consideradas normais as RNA são superiores ao modelo Logit Multinomial, mas nas situações de troca de marca extremas o modelo Logit Multinomial é superior; (4) as RNA têm uma capacidade previsional superior quando a variação do gosto do agregado familiar não é incluído; (5) as RNA são menos sensíveis ao número de observações a treinar, enquanto o modelo Logit Multinomial melhora a sua capacidade previsional com o número de observações na sua estimação. Genericamente, as RNA têm um comportamento superior ao modelo Logit Multinomial, mostrando uma melhor capacidade previsional. Têm também a vantagem de necessitar de menos preparação e reduzir o tempo de intervenção do analista, bem como requisitos computacionais inferiores. Contudo, embora teoricamente se possa proceder a uma análise de sensibilidade e obter elasticidades, estes dois itens são difíceis e morosos quando comparados com o modelo Logit Multinomial. Os autores concluem que os dois modelos devem ser aplicados de forma complementar.

Em Hruschka (2000) procede-se à mesma comparação, mas a RNA, embora tendo uma arquitectura unidireccional, tem diferenças substanciais. A sua organização deixa de ser em camadas e passa a incluir supra-conexões. A sua performance previsional é, contudo, superior ao modelo Logit Multinomial e revela ainda superioridade na identificação dos pontos de corte, bem como na determinação das elasticidades.

Em West et al. (1997) as RNA são comparadas com o modelo Logit Binário, procurando avaliar a performance

dos modelos no comportamento do consumidor na previsão da escolha com regras não compensatórias e quando a regra de decisão é desconhecida. Para esse efeito, foram realizados dois estudos: uma simulação numérica representativa de escolha com regras não compensatórias, utilizando números gerados aleatoriamente, e um inquérito para medir a percepção da imagem do ponto de compra e o seu relacionamento com o comportamento. Os resultados mostraram que as RNA tiveram uma capacidade previsionial superior, muito significativa nas regras não compensatórias e uma performance equivalente para a regra compensatória. Particularmente importante é a classificação quando são avaliados os erros do tipo I ou II, em que a RNA é mais equilibrada. No segundo teste, constituído por variáveis discretas nominais, a RNA revelou-se, uma vez mais, como modelo com melhores capacidades previsionais. Saliente-se, por último, que os autores descrevem um procedimento que permite aferir a variação da escolha em função da alteração de uma variável de entrada. Este facto é apontado muitas vezes como uma limitação ou desvantagem das RNA. No entanto, a substituição proporcional inerente ao modelo Logit não se verifica nas RNA, devido à utilização das funções activação sigmóides.

A demonstração que as RNA conseguem modelar o comportamento do consumidor, subdividindo-a numa primeira fase em que o decisor reduz o conjunto de escolha de alternativas a um conjunto de escolha de alternativas consideradas e, numa fase posterior, selecciona a alternativa, é descrita em Vroomen et al. (2004). Este faseamento da escolha é comparado pelos autores com uma síntese das fontes de informação mais relevantes que, numa perspectiva teórica é útil, mas numa perspectiva prática é questionável, pois esses conjuntos só podem ser conhecidos com preferências declaradas. Daí se reconhecer que o conjunto de escolha das alternativas consideradas é obtido estocasticamente. Os resultados obtidos mostraram que a RNA construída tem capacidades previsionais superiores ao modelo Logit Multinomial. Os autores também descrevem como é possível visualizar a interacção entre a variação dos valores das variáveis de entrada e a probabilidade de escolha das marcas.

Segundo Yip et al. (2001) as RNA só são superiores ao modelo Logit Multinomial quando a relação implícita entre as variáveis de entrada e as variáveis de saída é não monotónica. No caso da presença de uma relação monotónica, crescente ou decrescente, a performance do modelo Logit Multinomial é equivalente à das RNA. Para provar estas afirmações os autores executaram simulações numéricas tendo comprovado os pressupostos iniciais.

A aplicação dos AG na aprendizagem e a redução da arquitectura das RNA é proposta por Fish et al. (2002), mostrando que estas alterações melhoraram a performance da RNA e são mais eficientes do que o modelo Logit Multinomial. Assim, no seu estudo empírico, em que avaliam o comportamento do consumidor na compra de café entre seis alternativas, os autores começam por comparar uma RNA com arquitectura 42-11-6 e o modelo Logit Multinomial. Note-se que o número de variáveis de entrada corresponde a sete atributos vezes o número de alternativas. Nesta comparação o modelo Logit Multinomial foi superior, atribuindo-se-lhe essa superioridade à complexidade da RNA (quinhentas e quarenta e cinco conexões para mil quinhentos e sessenta e quatro pares de vectores). Em face destes resultados foi construída uma RNA com sete variáveis de entrada e uma variável de saída, impondo uma reformatação dos dados e obrigando os pares de vectores para cada escolha a serem apresentados seis vezes (igual ao número de alternativas). Designada por RNA com arquitectura reduzida a sua performance revelou-se superior ao modelo Logit Multinomial. Esta arquitectura reduzida foi alternativamente treinada com um algoritmo genético, tendo obtido resultados superiores a todos os outros métodos.

Embora a revisão das fontes de informação seja, na sua maioria, constituída por estudos comparativos entre RNA e os MED, existem alguns trabalhos que sugerem a integração destes dois modelos. Nesta perspectiva, um dos mais importantes é proposto por Bentz e Merunka (2000). Os autores constatarem o seguinte: (1) o modelo Logit Multinomial é aplicado primordialmente com uma função utilidade linear; (2) as RNA são modelos não paramétricos que não necessitam de pressupostos; (3) as RNA conseguem estabelecer relações complexas entre variáveis; (4) os coeficientes do modelo Logit Multinomial são interpretáveis; (5) as conexões de uma RNA não são interpretáveis, mas as RNA podem ser utilizadas na simulação de cenários e visualização de funções e na

possível identificação de efeitos não lineares. Analisando os itens descritos, verifica-se que a limitação da linearidade da utilidade no modelo Logit Multinomial não existe nas RNA. No comportamento do consumidor a limitação da linearidade da utilidade é uma desvantagem importante porque a modelação que inclua interação entre variáveis, pontos de corte ou outros efeitos não lineares obriga o analista a uma pesquisa extensa de combinações das variáveis, que pode resultar em fracasso e conseqüente enviesamento da especificação. As RNA permitem visualizar a relação entre as variáveis de entrada e as saídas dos neurónios, ajudando o analista a identificar relações não lineares. Para comprovar estes pressupostos, os autores utilizaram uma RNA com uma arquitectura unidireccional, conexões partilhadas para a mesma variável de entrada, função erro entrópica cruzada e função activação softmax para a camada de saída. A RNA foi treinada com o algoritmo BP e, posteriormente, foram criados gráficos representativos das variáveis de entrada e da probabilidade de escolha. Esses gráficos permitiram identificar relações não lineares que, uma vez incorporadas no modelo Logit Multinomial, melhoraram a sua performance. Em Fish e Blodgett (2002) é igualmente descrito um processo para representar as variáveis de entrada e de saída, permitindo identificar elasticidades e a importância de cada variável. Conclui-se, pois, que as RNA podem melhorar a performance dos MED, identificando interações entre variáveis e relações não lineares que, ao serem incluídas na função utilidade, aumentam as capacidades previsionais.

Este modelo foi posteriormente adaptado por Hruschka et al. (2001) e Hruschka et al. (2002), em que foi comprovada a sua superioridade sobre o modelo Logit Multinomial e a identificação de efeitos de interação entre variáveis, pontos de corte, efeitos de saturação e outros efeitos não lineares.

Em Papatla et al. (2002) são descritos modelos híbridos de RNA e MED que, na generalidade, têm uma capacidade previsionial superior à capacidade previsionial isolada de cada modelo. Estes modelos híbridos consistem em modelos de mistura de classes, resultantes de combinações lineares e não lineares dos modelos referidos. Os resultados obtidos indicam que as RNA, como modelo isolado, têm capacidades superiores ao modelo Logit Multinomial e os modelos de mistura são superiores a ambos.

3. APLICAÇÃO PRÁTICA

Na vertente das RNA são utilizados: o algoritmo BP e as suas variantes; os algoritmos construtivos; as RBF; os algoritmos de eliminação de parâmetros; os AG na optimização de RNA com arquitectura unidireccional. São igualmente analisados factores relevantes referidos nas fontes de informação como a inicialização dos valores das conexões, a aprendizagem incremental e batch, a regularização e a paragem antecipada da aprendizagem. As funções erro também são comparadas e avaliadas no desempenho dos algoritmos, nomeadamente a função erro média da soma do erro quadrado e a função erro entrópica. Nos MED foram seleccionados o modelo Logit Multinomial e o modelo Mixed Logit.

O objectivo primordial desta comparação residiu, essencialmente, na identificação do modelo com melhores capacidades previsionais para produtos com envolvimento fraco, pretendendo-se ainda obter outras ilações como a facilidade de implementação e a robustez. Para o efeito, os dois conjuntos de modelos, RNA e MED, foram calibrados independentemente e reconheceu-se a imperatividade de iniciar o processo pelas RNA porque, sendo um método não paramétrico, proporcionam poucas indicações da adequação das variáveis explicativas.

Para avaliar a performance dos MED e das RNA procedeu-se à aplicação dos diversos modelos numa amostra em que são adquiridos produtos com envolvimento fraco, nomeadamente detergentes líquidos. Esta amostra já foi utilizada em Chintagunta e Prasad (1998) e por Vroomen et al. (2004).

A amostra tem 3055 observações pertencentes a 400 agregados familiares. Das 3055 observações 398 são observações incompletas e, por essa razão, foram removidas da amostra. Assim, o total de observações disponíveis com informação completa é de 2657. O número de alternativas do conjunto de escolha é seis: (1)

Tide; (2) Wisk; (3) Eraplus; (4) Surf; (5) Solo; (6) All.

As variáveis incluídas na amostra, caracterizadoras da alternativa e do agregado familiar, são as seguintes: Gasto Noutros Produtos na mesma aquisição; Dimensão do Agregado Familiar; Volume Adquirido na Aquisição Precedente; Tempo Decorrido entre Aquisições; Preço de cada Alternativa; Promoção para cada Alternativa; Destaque para cada Alternativa. As variáveis de saída/dependentes são em número de seis e correspondem a cada alternativa. Para cada registo a variável escolhida assume o valor unitário, enquanto as outras variáveis assumem o valor nulo.

Antes de se proceder a qualquer simulação foi necessário realizar uma avaliação à amostra, quer para identificar possíveis outliers, quer para obter as percentagens de cada alternativa. Concluiu-se que a amostra não é balanceada e, por essa razão, segundo as fontes de informação, as alternativas com menos elementos tenderão a ser classificadas com menor precisão. A identificação de outliers ou extremos em amostras unidimensionais segue procedimentos simples, sendo, por isso, uma tarefa acessível. Mas em amostras multi-dimensionais essa identificação pode ser difícil. Alguns indicadores elementares como a média, o desvio padrão, o mínimo e o máximo são a primeira ajuda. Analisando os histogramas das variáveis concluiu-se, igualmente, a probabilidade de existirem outliers na amostra.

3.1. AVALIAÇÃO DAS REDES NEURONAIS ARTIFICIAIS

Perante a diversidade dos paradigmas das RNA, as simulações foram agrupadas em conjuntos, reflectindo as classes descritas nos capítulos precedentes. Assim, o primeiro conjunto de simulações incidiu nas RNA com arquitectura unidireccional e que foram treinadas com o algoritmo BP e variantes. Posteriormente, foram formados os conjuntos de simulações: algoritmos construtivos; algoritmos de eliminação de parâmetros; variantes das RNA RBF; optimização das RNA com AG.

Das fontes de informação ressaltou consistentemente que a identificação da arquitectura óptima da RNA para solucionar determinado problema é um processo “tentativa-erro”, em que se procura determinar o número de camadas interiores, o número de neurónios para cada camada interior e o tipo de função activação para cada camada interior. O número de neurónios das camadas de entrada e de saída é implicitamente definido pelo número de variáveis de entrada e de saída, obtendo-se para este caso vinte e duas variáveis de entrada e seis variáveis de saída. A função activação para as camadas interiores foi sempre a função activação tangente hiperbólica, porque proporciona uma convergência mais rápida do que a função activação logística.

Foram também definidas duas medidas de avaliação de desempenho das RNA: (1) média da soma do erro quadrado; (2) percentagem dos vectores classificados correctamente. A segunda medida de desempenho é calculada executando a sequência seguinte: submeter o par de vectores à RNA; identificar a classe que obteve maior probabilidade produzida pela RNA; se essa classe corresponde à mesma classe escolhida então incrementa-se o número de vectores classificados correctamente, caso contrário incrementa-se o número de vectores classificados incorrectamente.

Nesta fase a amostra foi dividida em três conjuntos: conjunto de aprendizagem (50% dos pares de vectores); conjunto de validação (25% dos pares de vectores); conjunto de teste (25% dos pares de vectores). Esta divisão é necessária para utilizar o método da paragem antecipada da aprendizagem, o qual assegura uma identificação mais rápida da arquitectura óptima. É ainda de referir dois aspectos importantes. O conjunto de teste é sempre disjunto do conjunto de aprendizagem e do conjunto de validação. Os conjuntos são formados aleatoriamente em cada simulação. Note-se que as medidas de desempenho foram sempre aplicadas aos conjuntos de aprendizagem e de teste. Mas a avaliação das medidas de desempenho finais foi feita sobre o conjunto de teste, pois são estas que indicam a capacidade previsionial da RNA. As medidas de desempenho apenas foram aplicadas ao conjunto de aprendizagem para assegurar que a divisão da amostra é equilibrada.

As primeiras simulações para determinar o número de neurónios adequado para a camada interior obtiveram

resultados muito fracos. As simulações começaram com 4 neurónios e terminaram com 50. Foi utilizado o algoritmo Levenberg-Marquardt (BPLM) com paragem antecipada da aprendizagem. Para a camada de saída foi escolhida a função activação logística e o número de épocas foi limitado a 25. Em cada simulação os valores das conexões seguiram a inicialização de Nguyen-Widrow, tendo-se previamente comprovado que essa inicialização proporciona uma convergência mais rápida. Às variáveis de entrada foi aplicado um pré-processamento para ficarem com média nula e desvio padrão unitário. Em presença de resultados medíocres procuraram-se identificar alternativas. O recurso às fontes de informação revelou-se, neste caso, bastante proveitoso com a indicação de variáveis a incluir na amostra, em específico a inclusão da variável Fidelização, indica a alternativa escolhida na aquisição precedente, e Fidelização_A, indica a penúltima alternativa escolhida pelo agregado familiar. A amostra ficou reduzida a 1937 observações.

Inevitavelmente, seguiu-se o processo iterativo para identificar o número de neurónios óptimo e as variáveis explicativas. Numa primeira fase executaram-se simulações para identificar as variáveis explicativas relevantes e, posteriormente, executaram-se simulações para determinar o número de neurónios óptimo. As simulações para identificar as variáveis explicativas foram efectuadas com uma RNA com 15 neurónios, tendo sido utilizado o algoritmo BPLM com paragem antecipada da aprendizagem. Para a camada de saída foi escolhida a função activação logística e o número de épocas foi limitado a 200. Em cada simulação os valores das conexões seguiram a inicialização de Nguyen-Widrow e as variáveis o pré-processamento referido atrás. De todas as simulações o erro médio mais pequeno foi obtido com as variáveis: Fidelização_A; Fidelização; Gasto Noutros Produtos na mesma aquisição; Dimensão do Agregado Familiar; Volume Adquirido na Aquisição Precedente; Tempo Decorrido entre Aquisições; Preço de Tide; Preço de Wisk; Preço de EraPlus; Preço de Surf; Preço de Solo; Preço de All. Após a identificação das variáveis explicativas qualquer RNA terá sempre 12 variáveis de entrada e 6 variáveis de saída.

Após se terem identificado as variáveis explicativas foram executadas novas simulações para determinar o número de neurónios óptimos. Essas simulações foram executadas com a amostra balanceada. Para cada arquitectura foram executadas 30 simulações e foi escolhida a arquitectura que produziu o erro médio menor. As RNA foram treinadas com o algoritmo BPLM com paragem antecipada da aprendizagem, função activação logística para a camada de saída e o número de épocas foi limitado a 200. Em cada simulação os valores das conexões seguiram a inicialização de Nguyen-Widrow e as variáveis o pré-processamento já referido atrás. O erro médio menor foi obtido para uma RNA com 13 neurónios na camada interior, tendo-se, por essa razão, a arquitectura final 12-13-6.

Após se ter identificado o número de neurónios para a camada interior procedeu-se a simulações para avaliar o comportamento e o desempenho das variantes do algoritmo BP. As variantes testadas são as mais representativas, incluindo algoritmos baseados em heurísticas, em métodos de segunda ordem e também com aprendizagem incremental. Na Tabela 1 estão listados os algoritmos testados com o respectivo número de épocas permitido para aprendizagem, bem como os valores das épocas verificados. Note-se que os parâmetros específicos de cada algoritmo foram devidamente calibrados.

Tabela 1 – Parâmetros e resultados da aprendizagem das variantes do algoritmo BP.

Variantes do algoritmo BP	Número máximo de Épocas	Épocas verificadas		
		Mínimo	Máximo	Média
Quasi-Newton BFGS	1500	12	152	65.17
Powell-Beale GC	1500	24	174	69.97
Fletcher-Powell GC	1500	8	134	45.93
Polak-Ribiere GC	1500	35	153	77.27
Gradiente (incremental)	5000	5000	5000	5000

CITIES IN COMPETITION

Variantes do algoritmo BP	Número máximo de Épocas	Épocas verificadas		
		Mínimo	Máximo	Média
Gradiente com coeficiente aprendizagem variável	5000	89	456	219.5
Gradiente com momentum	5000	5000	5000	5000
Gradiente com momentum e coef. aprend. variável	1500	124	315	195.6
BPLM	300	21	58	32.7
BPRPROP	5000	33	170	93.7
BPGCS	1500	28	165	82.83

A

Tabela 2 contém as medidas de desempenho para as variantes do algoritmo BP. De todas as variantes aquela que apresenta resultados melhores é o algoritmo BPLM. Esse facto verifica-se para todas as medidas efectuadas no conjunto de teste. O algoritmo tem o erro mínimo menor, o erro máximo menor e o erro médio menor e tem as percentagens de vectores classificadas correctamente mínima, máxima e média maiores. Os outros algoritmos de segunda ordem tiveram resultados aproximados entre si, evidenciando-se nos algoritmos com heurísticas o BPRPROP. Com um desempenho fraco sobressaem novamente os algoritmos BP com aprendizagem incremental e o com momentum.

Tabela 2 – Identificação da melhor variante do algoritmo BP.

Variantes do algoritmo BP	Erro do conjunto de teste			Percentagem de vectores classificados correctamente					
				Teste			Aprendizagem		
	Mínimo	Máximo	Média	Mín.	Max.	Méd.	Mín.	Max.	Méd.
Quasi-Newton BFGS	0.2871	1.1532	0.5085	21.54	72.31	55.71	16.91	81.39	61.78
Powell-Beale GC	0.2608	0.4639	0.3473	59.87	77.18	70.10	49.34	86.10	76.08
Fletcher-Powell GC	0.3430	1.0035	0.5833	23.85	71.03	52.71	22.46	81.46	61.35
Polak-Ribiere GC	0.2703	0.5998	0.3467	50.64	75.90	69.92	66.08	87.76	79.45
Gradiente (incremental)	0.4082	1.4672	0.8353	6.41	55.00	28.68	16.97	62.80	42.32
Gradiente com coeficiente aprendizagem variável	0.3015	0.5975	0.3713	30.64	71.67	60.55	29.49	79.87	64.55
Gradiente com momentum	0.5311	1.4427	0.8753	6.79	40.00	23.90	26.89	50.18	39.01
Gradiente com momentum e coeficiente aprendizagem variável	0.2767	0.5703	0.4118	30.90	74.36	55.86	36.13	83.25	61.81
BPLM	0.2492	0.3169	0.2706	70.90	78.08	76.06	62.66	88.94	84.94
BPRPROP	0.2640	0.3501	0.3058	63.85	76.03	70.04	74.69	83.81	80.33
BPGCS	0.2600	0.4974	0.3118	51.28	77.69	71.01	36.00	87.48	78.38

Após avaliar os resultados apresentados atrás pode concluir-se que o algoritmo BPLM obteve os melhores desempenhos para as medidas definidas e foi também o que demonstrou melhor robustez. A Tabela 3 contém as medidas de desempenho para três algoritmos com função erro entrópica.

Tabela 3 – Identificação da melhor variante do algoritmo BP com função erro entrópica.

Variantes do algoritmo BP	Erro do conjunto de teste			Percentagem de vectores classificados correctamente					
				Teste			Aprendizagem		
	Mínimo	Máximo	Média	Mín.	Max.	Méd.	Mín.	Max.	Méd.
BPGCS	0.4392	0.4533	0.4467	71.41	77.95	74.26	86.78	90.94	88.92
Quasi-Newton BFGS	0.4521	0.4729	0.4590	64.91	73.08	69.69	79.62	87.64	85.70
Gradiente (incremental)	0.4695	0.5442	0.5160	29.23	57.05	37.90	35.96	65.38	52.46

Os valores da Tabela 3 demonstram que a aplicação da função erro entrópica pode produzir resultados melhores do que a aplicação da função erro média da soma do erro quadrado para alguns algoritmos. Verifica-se, no entanto, que o erro mínimo é sempre maior com a função erro entrópica. A vantagem reside na obtenção de valores melhores para a percentagem média de vectores classificados correctamente no conjunto de teste.

Foram efectuadas simulações para avaliar o efeito da regularização. Assim, numa primeira fase avaliou-se o efeito da regularização mantendo o conjunto de teste (Tabela 4). Numa segunda fase avaliou-se o efeito da regularização sem o conjunto de validação. Todas as simulações foram executadas com o algoritmo BPLM, inicialização de Nguyen-Widrow e as variáveis de entrada com pré-processamento. O número de épocas foi limitado a 200.

Tabela 4 – Medidas de desempenho com regularização e paragem antecipada da aprendizagem.

Valor de regularização	Erro do conjunto de teste			Percentagem de vectores classificados correctamente					
				Teste			Aprendizagem		
	Mínimo	Máximo	Média	Mín.	Max.	Méd.	Mín.	Max.	Méd.
0.5	0.2599	0.6204	0.35992	53.59	78.46	71.05	71.04	91.48	85.90
0.6	0.2571	0.3568	0.30441	72.95	79.23	76.10	88.41	92.66	90.30
0.7	0.2748	0.4040	0.33151	69.23	78.72	74.00	74.27	92.00	88.30
0.8	0.2699	0.4122	0.33213	64.49	76.92	73.00	75.62	91.00	88.78
0.9	0.2693	0.7065	0.36282	44.23	78.08	71.68	58.85	93.24	86.61

A Tabela 4 mostra que o efeito da regularização com paragem antecipada da aprendizagem não melhora a performance das RNA. O valor da regularização que obteve os desempenhos melhores foi de 0.6.

As simulações com regularização e sem paragem antecipada da aprendizagem mostram os melhores resultados obtidos. Com excepção do valor 0.6, o erro mínimo obtido é sempre inferior ao algoritmo BPLM com paragem antecipada da aprendizagem e ao algoritmo BPLM com paragem antecipada da aprendizagem e regularização. Com o valor de regularização 0.5 obtém-se o menor erro médio. Em contrapartida, constata-se uma diminuição da robustez, verificando-se nalgumas situações uma convergência para um mínimo local, que se repercute em erros grandes. A Tabela 5 ilustra estes factos.

Tabela 5 – Medidas de desempenho apenas com regularização.

Valor de regularização	Erro do conjunto de teste			Percentagem de vectores classificados correctamente					
				Teste			Aprendizagem		
	Mínimo	Máximo	Média	Mín.	Max.	Méd.	Mín.	Max.	Méd.
0.5	0.2057	0.2901	0.26074	75.64	82.69	78.41	83.00	89.74	87.76
0.6	0.2658	1.2289	0.40483	22.69	78.46	69.00	26.17	90.00	76.01
0.7	0.2258	0.5176	0.28979	60.90	81.67	77.00	52.54	89.00	84.48
0.8	0.2396	0.3178	0.27099	74.10	80.13	78.00	84.29	89.00	87.29
0.9	0.2374	0.5648	0.29924	63.08	80.64	76.24	75.60	89.55	85.96

Saliente-se, por último, que todas as simulações necessitaram das 200 épocas e que o valor de regularização é obtido com recursos a várias experimentações, requerendo cálculo computacional adicional.

As RNA RBF foram também avaliadas. Das inúmeras variantes seleccionaram-se a RBF com arquitectura exacta, a RBF com heurística FS e a PNN. A RBF com arquitectura exacta tem um número de neurónios interiores igual ao número de pares de vectores de aprendizagem. À RBF com heurística FS foi imposto um máximo de 400 neurónios interiores ou um erro de aprendizagem de 0.01. A PNN segue as prescrições da RBF com arquitectura exacta.

Tabela 6 – Medidas de desempenho para RNA RBF.

Tipo de RNA	Erro do conjunto de teste			Percentagem de vectores classificados correctamente					
				Teste			Aprendizagem		
	Mínimo	Máximo	Média	Mín.	Max.	Méd.	Mín.	Max.	Méd.
RBF com FS	0.4992	0.5095	0.5044	57.44	64.23	60.73	65.49	69.35	67.86
RBF exacta	0.4391	0.4952	0.4520	62.56	72.69	68.19	100	100	100
PNN	0.3260	0.3690	0.3563	74.40	78.46	75.56	97.51	98.97	98.43

A PNN teve medidas de desempenho claramente superiores às outras RBF. No entanto a sua performance é inferior a RNA treinadas com o algoritmo BPLM e regularização ou a uma RNA treinada com o algoritmo BPLM e paragem antecipada da aprendizagem. Saliente-se, contudo, que a PNN obteve a média maior para a percentagem de vectores classificados correctamente.

Para os algoritmos construtivos foram seleccionados os algoritmos Cascade-Correlation e o Cascade-Correlation Sibling/Descendant. Para os dois algoritmos escolheu-se o algoritmo de aprendizagem BPRPROP, 64 neurónios candidatos, inicialização aleatória das conexões com valores entre [-1,1] e função activação logística para a camada de saída.

Tabela 7 – Medidas de desempenho para Cascade-Correlation e SDCC.

Tipo de RNA	Erro do conjunto de teste			Percentagem de vectores classificados correctamente					
				Teste			Aprendizagem		
	Mínimo	Máximo	Média	Mín.	Max.	Méd.	Mín.	Max.	Méd.
CC	0.4889	0.5230	0.5190	65.90	72.95	71.05	69.80	71.75	70.39
SDCC	0.4505	0.5037	0.4971	76.67	83.97	82.10	80.87	84.67	82.22

Os dois algoritmos tiveram erros médios grandes, distantes dos obtidos pelo algoritmo BPLM. No entanto, o algoritmo SDCC obteve uma boa média para a percentagem de vectores classificados correctamente.

A integração das RNA e dos AG foi elaborada num método Lamarckiano (GANN) em que o AG tem codificado os valores das conexões e optimiza-os durante várias gerações usando um procedimento de selecção dos melhores indivíduos. Neste modelo, o AG usa uma estratégia que garante que os melhores indivíduos encontrados durante sucessivas gerações são mantidos na população, seleccionando as melhores soluções entre parentes e descendentes. Esta estratégia assegura que as melhores soluções não são perdidas, não permite a repetição de indivíduos e proporciona uma convergência mais rápida. Posteriormente, o algoritmo BPLM treina a RNA com os valores das conexões optimizadas pelo AG.

O modelo funciona com uma sequência de etapas iterativas que se descrevem a seguir. O AG é inicializado com uma população de 500 indivíduos, em que cada indivíduo representa uma arquitectura unidireccional, e posteriormente evolui durante 1000 gerações. O AG é aplicado com adaptação com escalonamento linear, procedimento de selecção dos melhores indivíduos, operador genético de cruzamento com um ponto de corte e

mutação bit a bit, em que a probabilidade da mutação de cada bit é igual à inversa do número de bits do cromossoma. Cada conexão foi codificada com 29 bits e foi utilizada a codificação Gray. Depois da fase de evolução o modelo entra numa fase iterativa. Os cromossomas são descodificados em arquitecturas unidireccionais e treinadas com o algoritmo BPLM durante 10 épocas. Se nenhuma das RNA convergiu são de novo codificadas e o AG faz evoluir de novo a população durante 1000 gerações. O ciclo termina quando uma das RNA convergiu ou o número de ciclos atingiu o limite.

Tabela 8 – Medidas de desempenho para GANN.

Tipo de RNA	Erro do conjunto de teste			Percentagem de vectores classificados correctamente					
				Teste			Aprendizagem		
	Mínimo	Máximo	Média	Mín.	Max.	Méd.	Mín.	Max.	Méd.
GANN	0.1840	1.3168	0.238758	19.74	85.13	80.15	32.44	85.58	81.78

A GANN proposta obteve as medidas de desempenho melhores: erro mínimo e erro médio menores. Mas não é tão robusta, pois em pelo menos uma simulação obtiveram-se medidas de desempenho más.

Os algoritmos de eliminação de parâmetros foram o Optimal Brain Damage (OBD) e o Optimal Brain Surgeon (OBS). Estes algoritmos foram aplicados a RNA treinadas com o algoritmo BPLM. Os resultados obtidos mostram a superioridade do algoritmo OBS sobre o algoritmo OBD. Mostram também que a aplicação de algoritmos de eliminação de parâmetros aumenta o erro e, neste caso, verifica-se a diminuição da percentagem dos vectores classificados correctamente.

Tabela 9 – Medidas de desempenho para os algoritmos de eliminação de parâmetros.

Tipo de RNA	Erro do conjunto de teste			Percentagem de vectores classificados correctamente					
				Teste			Aprendizagem		
	Mínimo	Máximo	Média	Mín.	Max.	Méd.	Mín.	Max.	Méd.
OBD	0.3101	0.6711	0.5607	54.23	68.09	65.20	57.61	71.32	67.78
OBS	0.2989	0.6003	0.4604	55.54	70.22	67.70	60.64	73.31	68.75

O maior problema das RNA é a sua calibração. Sem um método para identificar as variáveis independentes relevantes e a posterior necessidade de identificar a arquitectura óptima, resta o recurso à execução de simulações. Este processo, algo ineficiente, é exigente em recursos computacionais. A acrescentar a esta situação estão as inúmeras variantes e paradigmas existentes e que poderão ter comportamentos adequados ou revelar-se pouco eficazes para o problema em estudo.

Verificou-se neste estudo que a informação adquirida anteriormente é crucial, permitindo melhorar significativamente a capacidade previsionial das RNA através da inclusão de novas variáveis independentes. Este factor revelou ser o mais importante durante todo o processo de calibração. A inclusão das variáveis Fidelização e Fidelização_A fez diminuir o erro médio e aumentar a percentagem média de vectores classificados correctamente.

A identificação das variáveis independentes explicativas seguiu um processo iterativo de eliminação progressiva de variáveis irrelevantes. Contudo, a identificação das variáveis independentes relevantes pode depender da arquitectura, isto é, do número de neurónios e o número de neurónios da camada interior adequado para a resolução pode depender do número de variáveis de entrada. Esta interdependência dificulta as duas tarefas. A identificação das variáveis independentes foi aferida através das execuções que produziram o erro médio menor. A mesma medida foi utilizada para a identificação do número de neurónios da camada interior, mas a proximidade dos resultados obtidos tornou mais difícil a sua escolha. Por outro lado, o número de neurónios que é sugerido por esta medida não maximiza as medidas de desempenho percentagem de vectores classificados

correctamente.

De todas as variantes de aprendizagem do algoritmo BP aquele que exibiu medidas de desempenho melhores foi o algoritmo BPLM, revelando ainda uma robustez maior. Seguiu-se-lhe o algoritmo BPRPROP com valores muito próximos. A aprendizagem incremental revelou-se uma alternativa inferior à aprendizagem batch. A função erro entrópica, sugerida frequentemente nas fontes de informação como a mais adequada para problemas de classificação, não obteve sempre resultados melhores do que a função erro soma do erro quadrado. A aprendizagem com regularização melhora os desempenhos dos algoritmos que são aplicados com paragem antecipada da aprendizagem. A aprendizagem com regularização e paragem antecipada da aprendizagem obteve resultados piores. Os paradigmas das RNA RBF, algoritmos construtivos e eliminação de parâmetros obtiveram resultados piores do que a melhor variante do algoritmo BP.

De todos os paradigmas avaliados aquele que obteve resultados melhores no estudo efectuado foi a GANN, mas não foi o paradigma mais robusto. Corresponde também ao algoritmo que exige mais recursos computacionais.

3.2. AVALIAÇÃO DOS MODELOS DE ESCOLHA DISCRETA

Os MED Logit Multinomial e Mixed Logit foram avaliados com a mesma amostra. Para garantir uma comparação correcta foram utilizadas as medidas de desempenho definidas previamente. Pela mesma razão os modelos foram aplicados na amostra inicial e, posteriormente, nas amostras com a inclusão das variáveis Fidelização e Fidelização_A. Iniciou-se a avaliação dos MED com o modelo Logit Multinomial, tal como é sugerido nas fontes de informação. Para a amostra inicial as medidas de performance foram fracas, piorando com a inclusão das variáveis Fidelização e Fidelização_A. Perante resultados tão fracos foi necessário proceder à averiguação de outras variáveis explicativas e mais adequadas ao tipo de especificação utilizada nos MED. Após várias tentativas e simulações foi definida a variável Fid de Guadagni e Little. Como a amostra em estudo não avalia a fidelização quanto à dimensão da alternativa, a variável Fid contempla apenas a fidelização à marca. Assim, a equação que define Fid é:

$$Fid_{j,t} = \gamma Fid_{j,t-1} + (1-\gamma)y_{j,t-1}$$

Em que γ é um parâmetro de atenuação, $y_{j,t-1}$ assume o valor unitário se a alternativa j foi a escolhida na aquisição $t-1$. A condição inicial para esta variável impõe que $\sum (Fid_{j,1}) = 1$. O valor atribuído a γ foi de 0.9.

O modelo Logit Multinomial foi estimado com 75% dos valores da amostra, sendo os 25% restantes guardados para avaliar a capacidade previsionial do modelo. Foram sempre executadas 30 simulações com conjuntos disjuntos e formados aleatoriamente.

A construção da função utilidade de cada alternativa segue um processo iterativo de eliminação de variáveis tendo como finalidade obter o \bar{p}^2 maior e, simultaneamente, obter um modelo com o menor número de parâmetros. A eliminação das variáveis requer algum cuidado, pois, por vezes, verificou-se que a eliminação de uma variável baseada apenas no rácio t diminui o \bar{p}^2 e afecta as outras variáveis, significando uma interacção entre as mesmas.

Na

Tabela 10 estão as medidas de desempenho obtidas para as 30 simulações. O modelo estimado exibe medidas de desempenho boas e demonstra uma boa capacidade previsional. A característica mais saliente é, no entanto, a sua robustez, pois foi obtido um intervalo de valores reduzido.

Tabela 10 – Medidas de desempenho para o Logit Multinomial com a variável Fid.

Erro do conjunto de teste			Porcentagem de vectores classificados correctamente					
			Teste			Aprendizagem		
Mínimo	Máximo	Média	Mínimo	Máximo	Média	Mínimo	Máximo	Média
0.2881	0.3099	0.2998	68.45	70.92	69.62	69.44	71.07	69.95

O modelo Mixed Logit foi avaliado apenas com a variável Fid. A inclusão das variáveis Fidelização e Fidelização_A revelou-se igualmente ineficaz, omitindo-se, por esta razão, esses resultados. O objectivo final é obter um modelo com as melhores capacidades previsionais. Sustentado neste pressuposto foram ignoradas as indicações sugeridas nas fontes de informação. Assim, a opção por coeficientes aleatórios ou por componentes do erro e a escolha das funções distribuição teve sempre como premissa a maximização da capacidade previsual do modelo, considerando as medidas de desempenho definidas. Como tal, a especificação seleccionada foi a de coeficientes aleatórios, escolhendo-se a distribuição triangular para o coeficiente do preço. As funções utilidade das alternativas são as mesmas que foram identificadas para o modelo Logit Multinomial. Os resultados obtidos para as 30 simulações estão representados na Tabela 11.

Tabela 11 – Medidas de desempenho para o Mixed Logit com a variável Fid.

Erro do conjunto de teste			Porcentagem de vectores classificados correctamente					
			Teste			Aprendizagem		
Mínimo	Máximo	Média	Mínimo	Máximo	Média	Mínimo	Máximo	Média
0.2788	0.3124	0.2954	68.90	71.28	70.21	69.19	70.11	69.64

O modelo Mixed Logit apresenta resultados superiores ao modelo Logit Multinomial, ainda que essa superioridade seja reduzida ou mesmo insignificante. O acréscimo computacional e a complexidade inerentes ao modelo Mixed Logit não justificam de todo a sua utilização para este problema.

7. CONCLUSÕES

Os subcapítulos anteriores contêm os resultados das simulações executadas com vários paradigmas das RNA e com os modelos Logit Multinomial e Mixed Logit. Para avaliar a capacidade previsual das RNA e dos MED foram definidas duas medidas de desempenho. Por razões óbvias, considerou-se prioritária a optimização da medida de desempenho média da soma do erro quadrado. Assim, considerando esta medida de desempenho, as RNA foram superiores com o algoritmo BPLM e paragem antecipada da aprendizagem, com o algoritmo BPLM e regularização e com a GANN. Os outros paradigmas da RNA obtiveram valores superiores aos modelos Mixed Logit e Logit Multinomial. A conclusão a retirar das afirmações anteriores é que, usando a medida de desempenho média da soma do erro quadrado, não se verificou uma superioridade das RNA sobre os MED e que também não se verificou uma superioridade dos MED sobre as RNA. Fazendo a mesma análise para a medida de desempenho média da percentagem de vectores classificados correctamente conclui-se que, para as três amostras, as RNA são inevitavelmente superiores aos MED.

Esta incongruência das duas medidas de desempenho revela que a superioridade de um conjunto de modelos não deve ser aferida apenas com uma medida de desempenho. A medida média da percentagem de vectores classificados correctamente é utilizada preferencialmente na área das RNA, enquanto a medida média da soma do erro quadrado é aplicada nos MED. É de salientar, no entanto, a pouca adequação da primeira medida para avaliar probabilidades.

Quase tão importante com a capacidade previsual é a avaliação da robustez dos dois conjuntos de modelos.

Neste aspecto, os MED são claramente superiores às RNA. As simulações esclarecem mesmo que os algoritmos de aprendizagem das RNA convergem prematuramente para mínimos locais e ficam “presos” nesses mínimos. Estas ocorrências têm influências nas medidas de desempenho das RNA, diminuindo acentuadamente as suas performances médias.

Resumindo as conclusões sobre as medidas de desempenho pode afirmar-se que: (1) não existe uma superioridade entre os dois conjuntos de modelos, considerando a avaliação com a medida de desempenho média da soma do erro quadrado; (2) as RNA são superiores aos MED, considerando a medida de desempenho média da percentagem de vectores classificados correctamente; (3) os MED são mais robustos.

Depois de avaliar e comparar as capacidades previsionais dos dois conjuntos de modelos, realçam-se os aspectos metodológicos subjacentes aos mesmos. A identificação das variáveis independentes é um processo fundamental em qualquer método que funcione com aprendizagem com supervisão. Neste aspecto os dois conjuntos de modelos são distintos. As RNA são modelos não paramétricos e, por essa razão, não têm estatísticas que indicam a relevância das variáveis independentes, enquanto os MED, como modelos paramétricos, oferecem essa facilidade. Isto significa que a identificação das variáveis independentes relevantes nas RNA é um processo mais moroso, exige recursos computacionais adicionais e é mais complexo. Genericamente, este processo inicia-se com todas as variáveis independentes e, progressivamente, tentam-se eliminar as variáveis não aumentando o valor da função erro. Contudo, a correlação e/ou interdependência entre variáveis dificultam a obtenção do conjunto de variáveis independentes relevantes. Os MED seguem um processo mais simples, mas a interdependência entre variáveis independentes também pode dificultar a identificação de quais são relevantes. Ainda sobre esta temática é interessante verificar que para a mesma amostra as RNA e os MED identificaram sempre conjuntos de variáveis independentes relevantes diferentes, depreendendo-se que os dois conjuntos de modelos para a mesma amostra incorporam informação diferente e estabelecem relações diferentes entre as variáveis independentes e dependentes. Acresce também referir que a importância de uma variável independente para um dos conjuntos de modelos não implica que essa variável assumam a mesma importância explicativa para o outro conjunto de modelos. Pode acontecer que a inclusão de uma variável proporcione um aumento significativo das medidas de desempenho de um dos conjuntos e a inclusão dessa mesma variável no outro conjunto provoque um decréscimo acentuado da performance desses modelos. Outro factor importante a considerar é a própria codificação da variável. Uma codificação pode proporcionar bons desempenhos para um dos conjuntos e obter maus desempenhos no outro conjunto de modelos.

As conclusões sobre a identificação das variáveis podem ser sintetizadas em: (1) o processo de identificação das variáveis independentes relevantes para as RNA é mais moroso, complexo e exigente em recursos computacionais quando comparado com o mesmo processo de identificação para os MED; (2) as RNA e os MED identificam conjuntos de variáveis independentes relevantes diferentes; (3) uma variável independente pode melhorar a performance de um dos conjuntos de modelos e piorar a performance do outro; (4) o tipo de codificação influencia a performance dos dois conjuntos de modelos.

Apesar de não se fornecer informação sobre tempos de execução, neste aspecto particular, os MED têm uma vantagem clara e incomparável. O modelo Logit Multinomial é o mais rápido de todos os modelos testados e a diferença residual para o modelo Mixed Logit não justifica a utilização deste último. É também de salientar que o modelo Mixed Logit é geralmente mais rápido do que os diversos paradigmas das RNA (com excepção das RNA RBF e PNN). A GANN é o paradigma mais exigente em recursos computacionais, requerendo 6 horas para uma simulação, muito superior aos 2 a 3 segundos para a estimação de um modelo Logit Multinomial.

Começando por avaliar os MED, o modelo Mixed Logit obteve para a medida de desempenho média da soma do erro quadrado valores inferiores ao modelo Logit Multinomial, embora essa diferença seja pouco significativa. Por outro lado, a especificação do modelo Mixed Logit é mais complexa e a sua estimação mais morosa. Por estas razões, considerando as amostras analisadas e a área específica de estudo, omitindo sem perda de rigor

científico as suas limitações, bem como o objectivo principal que é aferir a capacidade previsionial, o modelo Logit Multinomial obteve bons desempenhos. O modelo Mixed Logit, embora mais flexível e sem as limitações reconhecidas, não demonstrou uma capacidade previsionial muito superior ao modelo Logit Multinomial. O trabalho experimental desenvolvido comprova que as RNA não são sempre superiores aos MED e que estes também não suplantam sempre as RNA. Os testes realizados foram efectuados numa amostra de produtos referentes a aquisições com um envolvimento fraco. Para este tipo de aquisições a informação adquirida em transacções anteriores é fundamental para aferir comportamentos futuros, tornando desnecessária a inclusão de construtoras como as convicções, percepções, emoções e outras. Da mesma forma, os dados revelados foram suficientes para criar modelos com uma boa capacidade previsionial, evitando a inclusão de preferências declaradas.

BIBLIOGRAFIA

- AGRAWAL, D. E SCHORLING, C. (1996): "Market share forecasting: An empirical comparison of artificial neural networks and multinomial logit model", *Journal of Retailing*, Vol. 72, nº 4, pp. 383-407.
- BENTZ, Y. E MERUNKA, D. (2000): "Neural networks and the multinomial logit for brand choice model: A hybrid approach", *Journal of Forecasting*, Vol. 19, nº 3, pp. 117-200.
- CHINTAGUNTA, P. E PRASAD, A. (1998): "An empirical investigation of the "dynamic McFadden" model of purchase timing and brand choice: Implications for market structure", *Journal of Business & Economic Statistics*, Vol. 16, nº 1, pp. 2-12.
- FISH, K. E BLODGETT, J. (2002): "A visual method for determining variable importance in an artificial neural network: an empirical benchmark study", *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 11, nº 3, pp. 244-254.
- FISH, K.; JOHNSON, J.; DORSEY, R. E BLODGETT, J. (2002): "Using an Artificial Neural Network Trained with a Genetic Algorithm to Model Brand Share", *Journal of Business Research*, Vol. 57, nº 1, pp. 79-85.
- GUADAGNI, P. E LITTLE, D. (1983): "A Logit model of brand choice calibrated on scanner data", *Marketing Science*, Vol. 1, nº 2, pp. 203-238.
- HRUSCHKA, H. (2000): "An Artificial Neural Net Attraction Model (ANNAM) to analyze market share effects of marketing instruments". Vienna University of Economics and Business Administration.
- HRUSCHKA, H.; FETTES, W. E PROBST, M. (2001): "An Empirical Comparison of the Validity of a Neural Net Based Multinomial Logit Choice Model to Alternative Model Specifications". University of Regensburg, Germany.
- HRUSCHKA, H.; FETTES, W.; PROBST, M. E MIES, C. (2002): "A Flexible Brand Choice Model Based On Neural Net Methodology. A Comparison to the Linear Utility Multinomial Logit Model and Its Latent Class Extension", *OR Spectrum*, Vol. 24, nº 2, pp. 127-143.
- PAPATLA, P.; ZAHEDI, M. E ZEKIC-SUSAC, M. (2002): "Leveraging the strengths of choice models and neural networks: A multiproduct comparative analysis", *Decision Sciences*, Vol. 33, nº 3, pp. 433-468.
- VROOMEN, B.; FRANSES, P. E NIEROP, E. (2004): "Modeling consideration sets and brand choice using artificial neural networks", *European Journal of Operational Research*, Vol. 154, nº 1, pp. 206-217.
- WEST, P.; BROCKETT, P. E GOLDEN, L. (1997): "A comparative analysis of neural networks and statistical methods for predicting consumer choice", *Marketing Science*, Vol. 16, nº 4, pp. 370-391.
- YIP, Y.; KURUBARAHALLI, G. E SU, Y. (2001): "Influence of data structure in choice modeling: An empirical investigation using neural networks", *American Business Review*, Vol. 19, nº 1, pp. 67-75.