



Redes Neuronais Artificiais para estimar o preço da habitação em Portugal

Maria Cristina Canavarro Teixeira

Email: ccanavarro@esa.ipcb.pt; Fax: +351 272 339 901

Escola Superior Agrária

Instituto Politécnico de Castelo Branco, Portugal

José Maria Caridad y Ocerin

Email: ccjm@uco.es

Nuria Ceular Villamandos

Email: tdcevin@uco.es

Estadística, Investigación Operativa y Econometria

Universidad de Córdoba, Espanha

ÁREA TEMÁTICA: Desarrollo local y competitividad urbana

RESUMO:

Neste artigo, foi desenvolvido um modelo de Redes Neuronais Artificiais para estimar o preço de venda de um apartamento. Recorrendo à boa vontade de alguns Agentes Imobiliários, constitui-se uma base de dados com mais de trinta atributos, de apartamentos vendidos entre 2005 e 2009, numa cidade de Portugal. O agrupamento de alguns destes atributos, possibilitou a construção de índices de localização, de conforto, de anexos e de conservação que para além de ter a vantagem de incorporar várias características, reduz significativamente o número de variáveis explicativas no modelo. As variáveis explicativas utilizadas neste modelo, podem facilmente ser generalizadas para outras cidades do país. A utilização das Redes Neuronais Artificiais, para estimar o preço da habitação, pode ser bastante útil em Portugal, e é uma séria alternativa aos métodos econométricos.

PALAVRAS CHAVE: *Redes Neuronais Artificiais; Preço da habitação; Índices de atributos; Portugal*

1. Introdução

O estudo da evolução dos preços no mercado da habitação é importante por diversos motivos. Em primeiro lugar, porque pode constituir um ponto final a uma série de ligações que determinam o acesso da população ao mercado imobiliário (mediante uma comparação entre o preço médio e a renda familiar disponível). De facto, a razão fundamental que explica a dificuldade de acesso (compra) de habitação é o seu preço elevado.

Em segundo lugar, a habitação é um activo, que em comparação com outros bens de primeira necessidade, constitui o bem de primeira necessidade com preço mais elevado na nossa sociedade, sem dúvida, o mais importante (e o mais pesado) na carteira das famílias. O aumento do preço da habitação, supõe um efeito riqueza que pode ter consequências importantes sobre o equilíbrio macroeconómico, já que incrementos da riqueza, em teoria, darão lugar a aumentos de consumo das famílias e na procura conjunta.

Em terceiro lugar, o sector da habitação residencial apresenta períodos de forte expansão, seguidos de anos de recessão mais acentuada que qualquer outro sector económico, já que o preço da habitação e o ciclo económico estão estritamente relacionados. Por conseguinte, a sua evolução também tem consequências sobre o mercado de trabalho ligado à construção e sobre o mercado de materiais de construção.

Desta forma, o investimento no mercado de habitação não é o mesmo que investir num activo sem risco nenhum, pois a rentabilidade deste investimento, também pode chegar a ser negativa ou substancialmente inferior à gerada por outros activos. Existem múltiplos exemplos que ilustram esta afirmação: no Reino Unido, durante o *boom* do imobiliário no começo dos anos 70 (1970 – 1973) e até finais de oitenta (1986 – 1989) os preços experimentaram taxas de crescimento anuais, que em alguns casos, superaram os 20%. Este crescimento, foi seguido de uma contracção nos preços que caíram à roda dos 40% entre 1973 e 1977, enquanto entre 1989 e 1992 a caída foi de aproximadamente 30%; nos Estados Unidos o preço da habitação nova aumentou durante os anos 70 em 30%, enquanto na recessão do princípio de 90, os preços da costa oeste de EE.UU. chegaram a ter aumentos de 40%; outros casos mais recentes são o Japão e Hong Kong, ou no nosso enquadramento europeu mais próximo, o caso da Alemanha, Áustria ou Espanha.

Evidentemente que devemos assinalar a existência de mercados locais especiais, devido à sua localização geográfica. Em Portugal, por exemplo, existem consideráveis diferenças entre as diferentes regiões (interior e costa), grandes cidades e também dentro das mesmas.

A comparação do preço da habitação com outros bens de primeira necessidade constitui uma tarefa difícil. O preço da habitação, como o seu custo, apresenta uma grande heterogeneidade. O preço varia em função da localização, tamanho, tipo de habitação (moradias, apartamentos em bloco, etc.) qualidade de construção, etc. Além disso as características das habitações também variam com o tempo. Consequentemente a simples evolução do preço médio das habitações compradas e vendidas em cada período pode não ser, o indicador mais adequado para observar a evolução. As séries estatísticas disponíveis, recorrem, na maior parte dos casos, a correcções dos factores diferenciais mais óbvios, como o tamanho, utilizando a medição do preço médio por metro quadrado (embora esta correcção seja apenas superficial se tivermos em conta que a relação preço, área não é linear), isto é, do preço médio das habitações de um determinado tamanho (área).

Também se devia ter em conta a composição do agregado da habitação que se utiliza como representativo do parque total, isso é, o preço do metro quadrado obtido como resultado dependerá do número de habitações novas e usadas, assim como, se na nossa amostra existirem habitações para fins sociais.

Uma característica que o nosso mercado imobiliário apresenta, à semelhança de outros, é que os preços de transacção dos imóveis são variáveis não observadas, no sentido de que o que se dispõe é da informação prestada pelo comprador e vendedor sobre o preço do imóvel transaccionado.

Este facto assume especial importância no mercado imobiliário português dado o conhecido fenómeno de “fraude e evasão fiscal”, com o conseqüente risco de uma menor credibilidade da informação disponível sobre preços de transacção dos imóveis. Efectivamente, tenderá a haver uma sub estimacção dos preços de transacção declarados, de modo a suportar menores encargos fiscais com a aquisicção do imóvel (comprador) ou menos carga fiscal sobre os lucros ou mais-valias da venda (vendedor).

Por outro lado, relativamente aos valores da avaliacção pelas entidades de crédito com vista à concessão ao empréstimo bancário, poderá existir uma sob estimacção, devido à percentagem do valor de empréstimo ser inferior a 100%. Assim, com uma maior

avaliação do imóvel, o valor concedido no empréstimo cobre o montante do valor real de compra, neste caso inferior, ao valor da avaliação.

As dificuldades e limitações na disponibilidade e acesso a dados sobre transações e características dos imóveis são comuns a vários países.

2. Alguma revisão de literatura

A partir dos anos 90, a previsão com redes neuronais artificiais (RNA's) teve um tremendo avanço. Desde os trabalhos pioneiros de Borst em 1991, que os modelos com redes neuronais artificiais se têm tornado uma alternativa muito atractiva aos modelos econométricos tradicionais. A vantagem principal destas técnicas é a capacidade de lidar com as relações não lineares, e com formas funcionais inicialmente desconhecidas. A literatura mostra que há um misto de sucesso e insucesso com este método, provavelmente devido a diferentes variáveis de entrada e de diferentes condições de mercado.

Pouco tempo depois de Borst ter apresentado o seu trabalho, Do, Quang e Grudnitski em 1993, utilizam as RNA's para mostrar que o valor de uma casa desce significativamente com a sua idade, durante os primeiros 16 a 20 anos, resultado da deterioração física. A partir dessa idade, não só a diminuição do valor pára (devida ao tempo), mas também a casa começa a experimentar uma apreciação relacionada, em parte, pelo tamanho do seu lote. Neste trabalho foi usada informação proveniente de agentes imobiliárias, em San Diego, Califórnia, relativas a 242 moradias vendidas, no período de Janeiro de 1991 até Setembro de 1991.

Testes robustos de modelos de RNA's requerem a separação do conjunto de dados. É necessário definir um modelo de formação e um novo conjunto de dados para testar os modelos (Rossini, 1997). Esta metodologia foi aplicada para vários países, utilizando conjuntos de dados com as características específicas de cada local. Por exemplo em 1991, Borst usou as RNA's para conjuntos de dados de residências familiares na Nova Inglaterra, Do e Grudnitiski (1992) usaram dados de um serviço de listagem múltipla na Califórnia, enquanto Evans et al. (1993) trabalharam com habitação no Reino Unido. Worzala et al., (1995), Borst (1995), Borst et al., (1996), e McCluskey et al., (1997), usaram múltiplos conjuntos de aprendizagem, e compararam os resultados obtidos pelas RNA's e pelos modelos de Regressão Múltipla.

Segundo Borst (1995), a precisão das RNA's, torna-as rivais dos métodos de Regressão Linear Múltipla. O autor considera que estas, podem ser utilizadas na avaliação em massa, bem como num controlo de qualidade sobre os valores estimados por outros métodos. Em 1996, Borst e McCluskey atestam que as capacidades de previsão das RNA estão perfeitamente bem definidas através de estudos de investigação.

No ano seguinte, Rossini, fundamentado nos trabalhos dos seus antecessores, aplicou esta técnica para os dados do Sul Australiano. Para tal foram usados os dados relativos às vendas registadas pelo Department of Environment and Natural Resources (DENR) no Sul da Austrália, e acedidos através do sistema de vendas recuperadas UPmarket. O DENR recolhe os detalhes de todas as vendas ocorridas no Sul da Austrália, e torna-as acessíveis em formato digital. Um vasto conjunto de informação está disponível para cada propriedade incluindo detalhes da venda, valores de avaliação, informações sobre o local e características físicas no caso de se tratar de propriedades residenciais. No seu estudo, Rossini, utilizou três procedimentos para comparar os modelos de RNA's na estimação do valor do mercado imobiliário, com o modelo de regressão linear múltipla.

Neste seu trabalho, Rossini chegou à conclusão de o uso dos modelos de Regressão, é preferível em vez das RNA's, advertindo contudo que estes resultados não são completamente conclusivos (Rossini, 1997). Embora tenha chegado a esta conclusão (com o referido conjunto de dados), acredita que num futuro muito próximo, com o aumento das ferramentas computacionais, que as RNA's se venham a tornar uma poderosa ferramenta. Convicto desta certeza, em 1999, o autor apresenta conjuntamente com outro investigador outro trabalho com RNA's. Kershaw e Rossini (1999), usaram uma série de dados de casas para desenvolver *Constant Quality House Price Indices*, usando redes neuronais e técnicas econométricas. Neste trabalho ficou provado que as RNA's podem ser uma séria alternativa aos métodos econométricos.

Entretanto, Zhang, et al., (1998) apresentam o estado de arte da aplicação das RNA's em previsão. O objectivo destes autores foi o de sintetizar a investigação nesta área, o conhecimento profundo nas técnicas de modelar as RNA's, e sugerir a direcção futura de investigação. Há mais de dez anos, os investigadores ainda não tinham certezas sobre o efeito dos factores chave no desempenho das RNA's em previsão. Estes autores chegam à conclusão que as RNA's têm um desempenho satisfatório na previsão, em todos os campos. A sua adaptabilidade, a não linearidade e a sua capacidade de mapeamento de função arbitrária (*arbitrary function mapping ability*), são características únicas das RNA's que fazem com que estas sejam completamente

adequadas e úteis nas tarefas de previsão. As descobertas foram inconclusivas em relação a onde e quando as RNA's são melhores do que os métodos de previsão clássicos. Um número considerável de investigadores trabalhou no sentido de tentar chegar a uma conclusão. Há vários factores que podem afectar o desempenho das RNA's. Contudo, não existe uma investigação sistemática sobre este assunto (Zhang et al., 1998).

Mas se por um lado as redes neuronais são uma grande promessa no campo da investigação em previsão, por outro lado incorporam muita incerteza.

A análise da regressão linear múltipla, apresenta grandes dificuldades em lidar com a complexidade do mercado imobiliário, especialmente devido à correlação espacial e ao desconhecimento de forma funcional (González e Formoso, 2000). Estes autores consideram que de todos os atributos, o mais importante é a localização, relacionada com a fixação espacial do produto (imobilidade). O valor de localização está relacionado com a acessibilidade (oferta e qualidade de vias e meios de transporte) e com as características da vizinhança, ou seja, do uso do solo na envolvente próxima do imóvel. Medir estes efeitos é muito difícil, pois não são quantificáveis directamente, sendo medidos através de variáveis *proxy*, tais como a distância ao centro comercial/histórico da área urbana.

Em 2001, Nguyen e Cripps, recolheram um total de 3906 observações de residências uni-familiares vendidas, em Rutherford, através do site Multiple Listing Service for the Rutherford Country, Tennessee, para um período de 8 meses entre 1 Janeiro de 1993 e 30 Junho de 1994. Neste estudo, os autores conseguiram provar que o desempenho das RNA's é superior ao do MRLM, e dão uma explicação plausível para esse facto, enquanto que em investigações anteriores, estes resultados não eram assim tão consistentes (Nguyen e Cripps, 2001).

Um outro estudo com o mesmo objectivo, foi apresentado em 2004 por Limsombunchai, Gan e Lee. Para tal, foi usada uma amostra de 200 casas de Christchurch na Nova Zelândia, tendo sido aleatoriamente seleccionada através do site Harcourt em 2003. Os resultados também mostraram que as redes neuronais têm um melhor desempenho do que os modelos hedónicos, contudo este trabalho apresenta algumas limitações e uma grande desvantagem, que é a utilização do valor estimado da casa, e não do verdadeiro valor de mercado, isto é, os dados não correspondem a vendas efectivas. Este problema, referem os autores deve-se à grande dificuldade de obtenção dos dados reais de mercado.

O interesse em métodos não convencionais para estimação do preço do mercado imobiliário tem vindo a crescer, principalmente na última década. A maioria utiliza as Redes Neurais, e os resultados obtidos não são consensuais, mas no entanto é notório o crescimento do interesse nestes métodos (Worzala et al., 1995; McGreal et al., 1998; Nguyen et al., 2001; Connellan et al., 1998; Bee-Hua, 2000; Lokshina, et al., 2003).

Já em 2006, o trabalho de Zurada, Levitan e Guan, apresenta os resultados do uso de dois métodos não convencionais, a lógica fuzzy e o raciocínio baseado na memória (memory-based reasoning) na avaliação dos valores dos bens imóveis residenciais, tendo sido estes modelos aplicados a um conjunto de dados reais. Este artigo também compara os resultados obtidos com estes dois métodos e com os modelos de regressão múltipla e as RNA's. Métodos de tratamento prévio de dados, como a Análise de Componentes Principais e selecção de variáveis, também foram utilizados com o objectivo de melhorar os resultados finais. Contudo, os resultados indicam que nenhum dos dois modelos foi consistentemente superior, quando aplicado a este conjunto de dados, quando comparado com os resultados obtidos com as RNA's e a Regressão Linear Múltipla.

O trabalho mais recente, conhecido até à data, para previsão do preço de venda de casas construídas, desenvolveu um novo modelo baseado na lógica fuzzy (Kusan et al., 2010). Este sistema, considera o plano da cidade, a proximidade a edifícios culturais, médicos, desportivos e de educação, os transportes públicos e outros factores ambientais, assim como a crescente tecnologia associada à construção de novos edifícios. Estes factores foram tomados como variáveis de entrada no modelo construído, segundo o objectivo do trabalho. Toda a informação foi proveniente de agências imobiliárias. Os valores da previsão e os valores de venda das casas foram comparados, tendo sido obtida uma muito boa precisão de ajustamento (Kusan et al., 2010).

Num estudo anterior também para a Turquia, Selim, comparou um modelo hedónico com um modelo de RNA's para determinar o preço da habitação na Turquia (Selim, 2009). O autor usou os dados de 2004 constantes do Household Budget Survey Data for Turkey (Instituto de Estatística Turco) num total de 5741 observações englobando casas urbanas e rurais, contendo 46 variáveis caracterizadoras de cada habitação. As variáveis incluem o tipo de casa, a idade do edifício, o tipo de edifício, o número de quartos, a área, o sistema de aquecimento, entre outras características estruturais, e engloba também uma variável caracterizadora do local. Relembre-se que os dados são referentes a um país. No entanto factores ambientais não puderam ser considerados porque não

constam da base de dados. No fim do seu estudo provou que, dada a não linearidade dos modelos de regressão hedónicos, as RNA's podem ser uma melhor alternativa de modelação dos preços das casas na Turquia.

Em 2008, Noelia García, Matías Gámez e Esteban Alfaro, introduziram um sistema automático de avaliação de imóveis, que combina o uso de redes neuronais artificiais com um sistema de informação geográfico. A opção dos autores assenta no facto de que ambas as ferramentas já demonstraram a sua utilidade potencial no campo da investigação económica. Em 2002, Thurston declarou, que uma RNA ligada a um GIS pode ser usada para simular como é que o cérebro humana processa problemas de dados espaciais. Há imensas aplicações em que uma RNA acoplada ao GIS se tornou muito útil. Por exemplo, podemos mencionar, o uso do solo, a oceanografia, a floresta, o movimento dos consumidores, avaliação do ruído gerado por um aeroporto, entre outros. Thurston, conseguiu mostrar como alguns modelos de redes neuronais e um sistema de informação geográfica podem ser combinados, constituindo uma poderosa ferramenta na área da economia. García, et al, em 2008, afirmam que qualquer que seja a abordagem utilizada, a análise pode ser melhorada através da integração de um sistema de informação geográfica. No seu trabalho, estes investigadores usaram os modelos de RNA's, percepção multi-camada, funções de bases radiais e os mapas de Kohonen's. Os dois primeiros modelos são uma interessante alternativa aos modelos tradicionais de regressão, enquanto que os mapas de Kohonen's (SOM), está especialmente vocacionado para tarefas de *clustering*. Por isso os dois primeiros modelos foram usados para estimar o preço dos imóveis, enquanto que os mapas de Kohonen's foram usados para tarefas intermédias relacionadas com a estimação de valores faltantes para variadas variáveis qualitativas, tais como a qualidade da propriedade.

Neste interessante trabalho, os autores combinaram as RNA's com um sistema de informação geográfica, num sistema automático para estimação do preço de uma casa em Albacete, Espanha. Através de um simples *click*, em cima do mapa desta cidade, e após fornecer ao sistema as características da casa pretendida, o sistema devolve a estimativa para o preço da casa. Os resultados de desempenho dos modelos foram comparados, tendo-se obtido uma melhor precisão com as RNA's, na determinação da estimativa do preço total com um R^2 de 92% e um erro médio relativo de 5,56%. Os autores suspeitam que a razão para estes resultados seja a dimensão da amostra disponível (591), pequena para os requerimentos da rede de funções de base radiais. A análise de sensibilidade mostrou que a variável mais importante foi a distância ao

distrito central de negócios, com um declive negativo de acordo com o pressuposto monocêntrico. Outro resultado importante do estudo de Garcia et al., de 2008, foi a não linearidade existente na relação entre o efeito da idade no preço da casa. Neste sentido, vale a pena mencionar a capacidade dos modelos neuronais em detectar relações não lineares, que não podem ser detectadas através dos modelos mais tradicionais.

Outros estudos foram feitos para Espanha, utilizando diferentes áreas urbanas. Caridad et al, 2009, usaram uma amostra com mais de 10000 transacções, recolhidas em trabalhos anteriores entre 2002 e 2006, para a cidade de Córdoba, permitindo também fazer comparações temporais. No contexto que se vive actualmente, urge encontrar formas objectivas de determinar o verdadeiro valor das propriedades. Em Espanha, os preços dos imóveis são recolhidos pelo INE e pelos municípios, com fins fiscais, não estando concentrados na verdadeira avaliação da propriedade individual. Assim, foi através de inquéritos e de entrevistas aos mediadores imobiliários que os dados para este estudo foram recolhidos. Usaram uma RNA multi-camadas para modelar o preço, com seis variáveis de entrada escolhidas através de diversas técnicas não discriminadas pelos autores. A variável explicativa do preço mais importante foi a área, seguida do índice de localização, e das despesas com o condomínio. Os anos do edifício, o índice de extras (medido através da existência de arrecadação e garagem) e o índice de qualidade (janelas, chão, mobílias e cozinha), demonstraram ter menos peso, no entanto se retiradas do modelo, o resultado final é menos bom. Caridad et al, 2009, consideram que o uso das RNA's é mais flexível do que os modelos clássicos econométricos, quando um conjunto de dados suficiente está disponível.

Mas apesar da contribuição desses modelos, os resultados não são unânimes sobre a existência de modelos capazes de prever a variação do valor dos imóveis com o tempo. Como tal, há uma necessidade de modelos práticos e automatizados que ajudem a alcançar este importante objectivo (Khalafallah, 2008).

Resumindo, desde os anos 90 em que as Redes Neuronais Artificiais se começaram a aplicar na área do imobiliário, tem-se assistido ao surgimento de variados modelos de valorização do mercado imobiliário, em diversas regiões do planeta. É notória a procura crescente ao longo do tempo, de novos e melhores algoritmos de funcionamento das redes neuronais.

São também muitos os estudos que estabelecem uma comparação entre os sistemas de inteligência artificial e os métodos tradicionais de avaliação de imóveis, especialmente com a Regressão Múltipla. Para o efeito, é geralmente calculada a percentagem de erro

de um sistema de IA e de outro de Regressão Múltipla aplicando-os a uma amostra representativa do mercado para a qual se conhece o preço de venda dos imóveis. As vantagens que os sistemas de IA trouxeram relativamente aos métodos tradicionais podem resumir-se basicamente em dois:

- Os sistemas de IA, apresentam nas provas, erros médios que se situam entre 5 e 10%, enquanto os modelos de Regressão Múltipla têm erros entre os 10 e os 15%. Há no entanto que ressaltar que em algumas experiências os resultados das experiências de ambos são semelhantes, quando se trata de amostras homogéneas (Couto, 2007).
- A segunda vantagem de um sistema de IA é a sua capacidade para estimar o valor das propriedades que apresentam características significativamente diferentes das que estão nas proximidades (valores extremos ou *outliers*), dado que este tipo de sistemas submete as amostras a processos matemáticos muito mais complexos que o modelo de Regressão Múltipla. Por outro lado, em determinados estudos, os sistemas de IA apresentam dificuldades em estimar com precisão os valores das propriedades com características especiais, *outliers*.

Os sistemas de IA já funcionam em Espanha em determinadas áreas, como por exemplo no sistema desenvolvido pela Agência Tributária para detecção da fraude e evasão fiscal, no imposto sobre o valor acrescentado, IVA.

Em Espanha, e dentro da valorização do mercado imobiliário, podemos destacar as contribuições de Caridad e Ceular (2001), García Rubio (2004), Gallego (2004) e Lara (2005), com aplicação às cidades de Córdoba, Albacete, Madrid e Jaén respectivamente. Actualmente, a Direcção Geral do Cadastro está a desenvolver um projecto para elaborar a estimação do valor de cada imóvel em preços de mercado, com o fim de perseguir a fraude imobiliária utilizando redes neuronais.

Em qualquer aplicação econométrica à realidade portuguesa, a obtenção de dados fidedignos representa parte substancial do trabalho a realizar. No caso concreto do mercado imobiliário habitacional as dificuldades encontradas são acrescidas, devido ao facto de não existir uma série temporal para os preços de “venda” da habitação, suficientemente longa, e que contemple os diferentes atributos do bem residencial (Carvalho, 1995).

Paula Couto, na sua tese de Doutoramento, faz entre outras, uma aplicação com redes neuronais a uma base de dados recolhida pelo INE com o apoio do software JavaNNS. Neste trabalho, não é apresentada qualquer interpretação da RNA, assim como da sua

capacidade de estimação. No entanto, a autora deste trabalho acaba por concluir que o modelo de avaliação de apartamentos, obtido por regressão linear múltipla, corresponde a um bom modelo para avaliação em massa para Portugal continental, com vista à obtenção dos respectivos valores patrimoniais tributáveis (Couto, 2007).

Este é no entanto, o único trabalho que encontramos até à presente data, onde existe uma aplicação com redes neuronais, a dados sobre avaliação de imóveis em Portugal. Julgamos que a não existência deste tipo de estudos no nosso país, se deve principalmente à dificuldade de acesso aos dados. A necessidade de uma grande quantidade de informação relativa às características dos imóveis, assim como o valor de venda dos mesmos, torna quase impraticável este tipo de análise no nosso país.

3. A amostra e a constituição de índices

Castelo Branco, está situada na zona centro de Portugal, na região raiana da Beira Interior profundamente esvaziada, e com os sectores agrícolas tradicionais em crise. A cidade demonstrou, apesar disso, um dinamismo apreciável, fruto do investimento industrial e da dotação de equipamentos e serviços de âmbito regional.

Nos Censos de 2001, os números apontavam para 31 mil habitantes na Freguesia de Castelo Branco registando-se uma variação positiva de aproximadamente 15% face aos censos anteriores. Relativamente aos alojamentos, Castelo Branco registava em 2001, 16607 alojamentos familiares clássicos e 93 de outro tipo de alojamento, dos quais 44 eram barracas. Dos 16607 alojamentos familiares clássicos, 67% eram de residência habitual, 24% de residência secundária e apenas 2% estavam vagos para venda (INE, 2001).

A base de dados, foi constituída com a boa vontade de alguns Agentes Imobiliários da cidade, e a informação diz respeito a mais de 200 apartamentos vendidos na cidade de Castelo Branco, entre 2005 e 2009, através da intervenção destes agentes. O facto de os apartamentos terem sido efectivamente vendidos, permite-nos o uso da palavra preço, uma vez que a venda ocorreu com um determinado valor conhecido.

Aproximadamente metade dos apartamentos amostrados são novos, tendo-se registado para cada um deles, o seu verdadeiro valor de transacção, isto é, o preço, assim como um vasto conjunto de características como se pode ver no quadro seguinte (Quadro 1).

Quadro 1. Definição das variáveis

| Variável | Definição |
|-------------------|---|
| t | 1 a 5, consoante o ano de venda seja 2005 a 2009 respectivamente |
| Estado | 0 se é novo, 1 se é usado |
| m ² | área útil habitável |
| Nºasso | nº de assoalhadas |
| Nº WC | nº total de casas de banho |
| Varanda | 1 se tem varanda; 0 caso contrário |
| Lareira | 1 se tem lareira; 0 caso contrário |
| Ar Condicionado | 1 se tem pré-instalação; 2 se tem aparelhos; 3 se é central; 0 c.c. |
| Aquec_central | 1 se tem aquecimento central; 0 caso contrário |
| Janelas | 2 se estão em bom estado; 1 estado regular e 0 em mau estado |
| Electrodomésticos | 1 se tem electrodomésticos na cozinha; 0 caso contrário |
| Arrecadação | 1 se tem arrecadação; 0 caso contrário |
| Garagem | 1 se tem lugar estacionamento; 2 se tem garagem individual; 0 c.c. |
| Condomínio | Valor de prestação mensal de condomínio, em euros |
| Preço | Preço de venda, em euros |
| Andares | Nº de andares do prédio |
| Elevador | 1 se tem elevador; 0 c.c. |
| Videovigilância | 1 se tem sistema de vídeo-vigilância; 0 c.c. |
| Localização | Zona da cidade |

Nota: Foram recolhidas outras variáveis como o número de casas de banho com banheira e sem banheira; armários embutidos; estores eléctricos; vidros duplos; soalho; móveis de cozinha; gás canalizado; pintura interior; hidromassagem; antiguidade; estado de conservação do prédio e rampa de acesso ao prédio.

Dentro das variáveis recolhidas, temos as numéricas ou quantitativas e as não numéricas e qualitativas. Às qualitativas, foi atribuído um valor numérico para que pudessem ser tratadas e analisadas de forma quantitativa. Por exemplo, se estivermos a contabilizar a existência ou não de garagem, usamos os valores 1 e 0, respectivamente. Para além disso, a existência de um grande número de atributos qualitativos na base de dados, levou-nos à necessidade de agrupamento dessas variáveis, permitindo por um lado a utilização de mais informação com menos variáveis, e por outro, a possibilidade de serem tratadas de forma quantitativa. Podemos encontrar na literatura autores que defendem o recurso a estes índices (Richardson, 1973; Saura, 1995; Jaén e Molina, 1995; Caridad e Ceular, 2001; Tabales, 2007). Por outro lado, a formação destes índices tem algum grau de subjectividade ao atribuir um valor numérico a um bem qualitativo. Assim, para que a constituição dos índices e dos seus valores fosse o mais assertivo possível, foram feitas várias validações, nomeadamente junto dos Agentes Imobiliários.

Constituíram-se cinco índices, englobando as características qualitativas amostradas, nomeadamente um índice de Conforto, de Anexos, de Conservação, um Interno e um Externo. O índice de Conforto, é formado pela existência de varanda, ar condicionado, aquecimento central, lareira, vidros duplos e estores eléctricos. O índice de Anexos é relativo à arrecadação e à garagem que pode ser individual, ou lugar de estacionamento.

O índice de Conservação reflecte o estado geral do apartamento, bem como o estado da pintura, e o estado geral das janelas. O índice Interno reflecte a existência de armários embutidos, soalho, móveis de cozinha, electrodomésticos na cozinha e gás canalizado e o índice Externo está relacionado com as características do prédio onde se encontra o apartamento, e engloba a existência de elevador, o estado geral de conservação do prédio, a existência de rampa de acesso, e o sistema de vídeo vigilância.

Estes índices variam todos entre 0 e 1, para que sejam o mais homogéneo possível e para que todos tenham à priori a mesma importância relativa. Se o valor do índice se aproxima do um, então significa que as variáveis que o compõem se encontram no óptimo, se estiver perto de zero significa que as variáveis que o compõem se encontram em situação desfavorável.

4. Resultados da estimação da rede neuronal artificial

Apresentamos de seguida os resultados obtidos na estimação da rede neuronal¹. No Quadro 2, podemos constatar que o erro relativo do conjunto de treino é quase o dobro do obtido no conjunto de teste.

Quadro 2: Resultado da rede neuronal em termos de precisão

| Model Summary | | |
|---------------|----------------------|--|
| Training | Sum of Squares Error | 18,393 |
| | Relative Error | ,206 |
| | Stopping Rule Used | 1 consecutive step(s) with no decrease in error ^a |
| | Training Time | 0:00:00.281 |
| Testing | Sum of Squares Error | 2,960 |
| | Relative Error | ,114 |

Dependent Variable: Preço_inflação

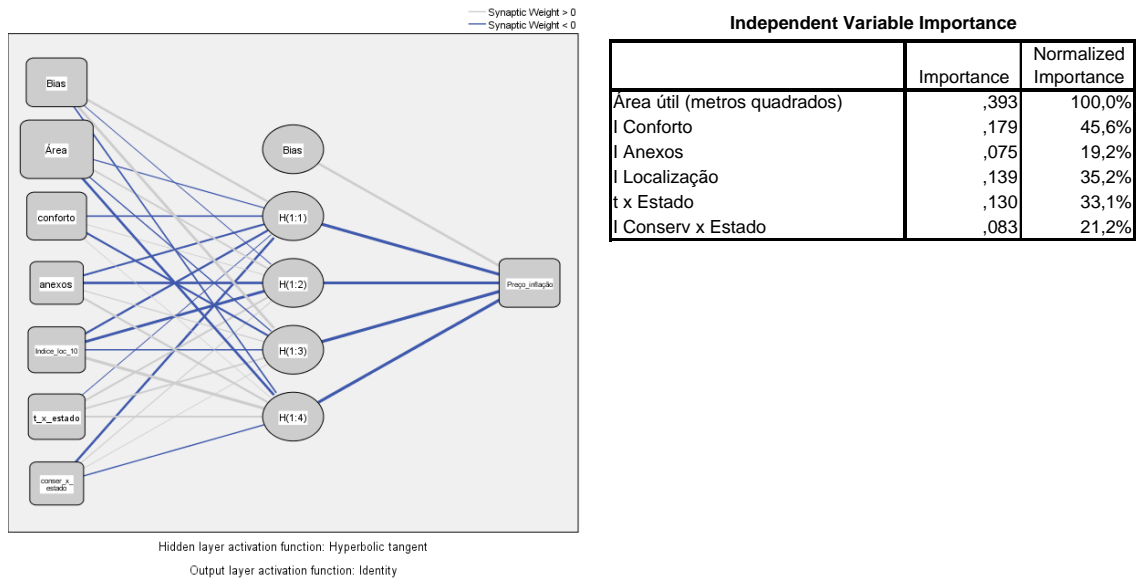
a. Error computations are based on the testing sample.

Fonte: *Output* do SPSS v.17

Na figura 1, podemos ver a rede neuronal estimada, a identificação do sinal dos parâmetros (pesos) calculados e a importância relativa das variáveis explicativas do preço estimado pela rede neuronal. No quadro 3, são apresentados os pesos das ligações entre os neurónios das várias camadas.

¹ A rede neuronal foi estimada com o módulo de redes neuronais do SPSS v.17; mais informações podem ser obtidas em www.spss.com.

Figura 1: Desenho da rede neuronal e importância relativa das variáveis



Fonte: *Output* do SPSS v.17

Quadro 3: Parâmetros estimados

| | | Predicted | | | | Preço_inflação |
|----------------|-----------------|----------------|--------|--------|--------|----------------|
| | | Hidden Layer 1 | | | | |
| Predictor | | H(1:1) | H(1:2) | H(1:3) | H(1:4) | |
| Input Layer | (Bias) | ,465 | -,110 | ,690 | -,232 | |
| | Área | -,192 | ,245 | -,199 | -,667 | |
| | conforto | -,210 | ,110 | -,375 | ,078 | |
| | anexos | -,365 | -,684 | ,184 | ,438 | |
| | Indice_loc_10 | -,457 | -,758 | -,207 | ,795 | |
| | t_x_estado | -,157 | ,348 | ,282 | ,286 | |
| | conser_x_estado | -,492 | ,164 | ,109 | -,200 | |
| Hidden Layer 1 | (Bias) | | | | | ,547 |
| | H(1:1) | | | | | -,841 |
| | H(1:2) | | | | | -,755 |
| | H(1:3) | | | | | -,948 |
| | H(1:4) | | | | | -,931 |

Fonte: *Output* do SPSS v.17

5. Conclusões

Podemos separar as conclusões em duas vertentes. Uma relacionada com a estimação do preço da habitação com RNA's, e a outra, com a importância das variáveis explicativas do preço da habitação em Castelo Branco

Começamos por analisar a capacidade da estimação da RNA obtida. O erro relativo do conjunto de teste obtido pelo SPSS (Quadro 2), permite-nos calcular a eficiência da rede (Pulido-Calvo et al., 2007). Neste caso, obteve-se uma eficiência aproximada de 89%.

Adicionalmente, foram calculadas algumas medidas habituais para avaliar o erro, e por conseguinte a precisão da rede neuronal ajustada. No quadro seguinte podemos ver esses resultados.

Quadro 4: Medidas de erro/precisão da Rede Neuronal Artificial

| SQE | MSE | RMSE | CV | MAPE | MAD | R ² |
|------------------|---------------|------------|--------|-------|------------|----------------|
| 4.013.547.783,24 | 95.560.661,51 | 9.775,51 € | 10,62% | 8,77% | 7.943,61 € | 88,97% |

Legenda: SQE – Soma dos quadrados do erro; MSE – Média da soma dos quadrados do erro; RMSE – Desvio padrão do erro, \sqrt{MSE} ; CV – Coeficiente de variação $\left(\frac{RMSE}{\bar{y}}\right) \times 100\%$; MAPE – Erro médio absoluto relativo; MAD – Erro médio absoluto e R² – Coeficiente de determinação.

Como podemos verificar pelos valores das diferentes medidas de erro, o processo de estimação dos preços através da rede neuronal, tem resultados bastante satisfatórios na estimação do preço de um apartamento em Castelo Branco. Inclusive, comparando estes resultados, com os obtidos pelo modelo de preços hedónicos, constatamos a superioridade dos resultados obtidos com a rede neuronal (Quadro seguinte).

Quadro 5: Medidas de erro/precisão do Modelo de Preços Hedónicos

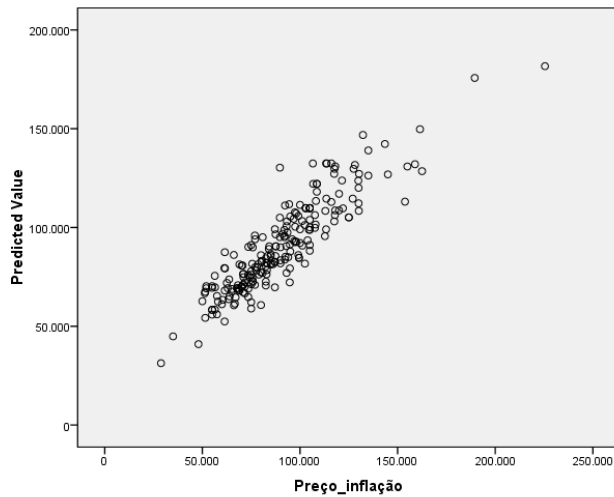
| SQE | MSE | RMSE | CV | MAPE | MAD | R ² |
|-------------------|----------------|------------|--------|--------|-----------|----------------|
| 36.053.024.942,28 | 162.401.013,25 | 12.743,67€ | 14,09% | 10,97% | 9.515,31€ | 77,99% |

Na rede neuronal, 54% dos erros são negativos, o que significa que no geral, há uma ligeira tendência para a rede sobrestimar o preço do apartamento, sendo o erro médio observado neste caso de -8.191,50 €

A média dos preços (observados) dos apartamentos que foram sobrestimados pela rede é de 82.189,91 € enquanto a média dos preços dos apartamentos subavaliados pela rede, é de 100.021,12 € Isto quer dizer, que a rede tem tendência para sobrestimar os apartamentos mais baratos, e de subavaliar os apartamentos mais caros. Isto é:

- Para os apartamentos mais baratos, isto é, com preço de venda observado inferior a 45 mil euros, a rede neuronal sobreavalia-os;
- Em relação aos apartamentos mais caros, isto é, cujo preço de venda observado foi superior a 140 mil euros, a rede neuronal subavalia-os.

Figura 3: Gráfico de dispersão do preço observado vs o preço estimado pela RNA



Fonte: *Output* do SPSS v.17

No gráfico da figura anterior (Figura 3) podemos comparar os preços de venda reais dos apartamentos, com os preços de venda estimados pela rede neuronal. O *insight* gráfico permite-nos concluir que o grau de ajuste da RNA é bastante bom, porque a maioria das observações está muito próxima da bissetriz do primeiro quadrante.

Analisando agora a outra vertente, relativa à importância relativa das variáveis utilizadas na estimação do preço de venda dos apartamentos na cidade de Castelo Branco, concluímos que a variável que mais contribui para explicar o preço total do apartamento é a área útil, com uma importância relativa de 39,3%. Note-se inclusive, que em todas as redes que foram estimadas na busca do “melhor modelo”, esta variável apresentou-se sempre como a mais importante na explicação do preço da habitação. De seguida, surge o índice de Conforto, com uma importância relativa de 17,9%, seguindo-se o Índice de localização e a interação entre o período de venda e o estado do apartamento. Com valores inferiores de importância relativa, temos a interação entre o estado da casa e o índice de conservação da mesma (para o caso dos usados) e por fim, o índice de anexos com percentagens de 8,3% e 7,5% respectivamente (como se pode ver nos resultados da Figura 1).

Julgamos que estas variáveis, pela quantidade de informação que incorporam, podem ser generalizáveis a outras cidades de Portugal, com as devidas adaptações, próprias de cada zona do país.

Agradecimentos

Aos responsáveis das Agências Imobiliárias GRADUZ, LING, IMOFACTOR e SGH de Castelo Branco, os nossos sinceros agradecimentos pela disponibilização de toda a informação fornecida, que tornou este estudo possível.

Bibliografia

Bee-Hua, B. (2000): “Evaluating the performance of combining neural networks and genetic algorithms to forecast construction demand: the case of the Singapore residential sector”, *Construction Management and Economics*, nº 2, p.209-218.

Borst, R. A. (1991): “Artificial neural networks: The next modelling/calibration technology for the assessment community?”, *Property Tax Journal*, IAAO, nº1, p.69-94.

Borst, R.A. e McCluskey (1996): “The Role of Artificial Neural Networks in the Mass Appraisal of Real Estate”, *Paper presented to the Third European Real Estate Society Conference*, Belfast, June 26-28.

Borst, R.A. (1995): “Artificial neural networks in mass appraisal”, *Journal of Property Tax Assessment & Administration*, nº2, p.5-15.

Caridad y Ocerin, J.M., Núñez Tabales, J., Ceular Villamandos, N. e Millán Vázquez, G. (2009): *27th International Conference. Mathematical Methods in Economics 2009*, Czech Republic.

Caridad, J. M. e Ceular, N. (2001): “Un análisis del mercado de la vivienda a través de Sistemas de Redes Neuronales”, *Revista de Estudios de Economía Aplicada*, nº18, p.67-81.

Carvalho, P. F. G. (1995): *O mercado de habitação em Portugal, Tese de Mestrado, Universidade de Coimbra. Faculdade de Economia, Portugal.*

Connellan, O. e H. James (1998): “Estimated realization price by neural networks: forecasting commercial property values”, *Journal of Property Valuation & Investment*, nº1, p.71-86.

Couto, P. (2007): *Avaliação Patrimonial de Imóveis para Habitação, Tese de Doutoramento, Laboratório Nacional de Engenharia Civil – Lisboa, Portugal.*

Do, A.Q. e Grudnitski, G. (1993): “A neural network analysis of the effect of age on housing values”, *J. Real Estate Res.*, nº2, p.253-264.

Evans A., James H. e Collins A. (1993): “Artificial Neural Networks: an Application to Residential Valuation in the UK”, *Journal of Property Valuation & Investment*, nº11, p.195-204.

Gallego Mora-Esperanza, J. (2004): "La inteligencia artificial aplicada a la valoración de inmuebles. Un ejemplo para valorar Madrid", *Revista CT/Catastro*, nº50, p.51-67.

García Rubio, N. (2004): *Desarrollo y aplicación de redes neuronales artificiales al mercado inmobiliario: aplicación a la ciudad de Albacete, Tesis Doctoral, Universidad de Castilla – La Mancha.*

González, M. A. S. e Formoso, C. T. (2000): “Análise conceitual das dificuldades na determinação de modelos de formação de preços através de análise de regressão”, *Centro de Engenharia Civil da Universidade do Minho, Revista de Engenharia Civil*, nº8, p.65-75.

Zhang G., Patuwo B. E. e Hu M. Y. (1998): “Forecasting with artificial neural networks: The state of the art”, *International Journal of Forecasting*, nº14, p.35–62.

INE (2001): *Censos 1991 e 2001, resultados definitivos, Instituto Nacional de Estatística de Portugal.*

Jaén, M. e Molina, A. (1995): *Modelos econométricos de tenencia y demanda de vivienda, Editorial Universidad de Almería.*

Kershaw, P. e Rossini, P. (1999): “Using Neural Networks to Estimate Constant Quality House Price Indices”, *Fifth Annual Pacific-Rim Real Estate Society Conference*, Kuala Lumpur, Malaysia.

Khalafallah, A. (2008): “Neural Network Based Model for Predicting Housing Market Performance”, *Tsinghua Science and Technology*, Number S1, p.325-328.

Kusan H., Aytekin O. e Ozdemir, Í. (2010): “The use of fuzzy logic in predicting house selling price”, *Expert Systems with Applications*, nº37, p.1808-1813.

Lara Cabeza, J. (2005): “Aplicación de las redes neuronales artificiales al campo de la valoración inmobiliária”, *Revista Mapping*, nº 104, p.64-71.

Limsombunchai, V., Gan C. e Lee, M. (2004): “House Price Predication: Hedonic Model vs. Artificial Neural Network”, *American Journal of Applied Science*, nº3, p.193-201.

- Lokshina, Hammerslag, e Insinga (2003): “Applications of artificial intelligence methods for real estate valuation and decision support”, *In Hawaii international conference on business*, Honolulu.
- McCluskey, W. e Borst, R. (1997): “An evaluation of MRA, comparable sale analysis, and ANNs for the mass appraisal of residential properties in North Ireland”, *Assesment Journal*, nº1, p.47-55.
- McGreal, S., A. Adair, D. McBurney, e D. Patterson (1998): “Neural Networks: the prediction of residential values”, *Journal of Property Valuation & Investment*, nº1, p.57-70.
- Nguyen, N. e Cripps, Al. (2001): “Predicting housing value: a comparison of multiple regression analysis and artificial neural networks”, *The Journal of Real Estate Research*, nº3, p.313-336.
- García N., Gámez M. e Alfaro E. (2008): “ANN+GIS: An automated system for property valuation”, *Neurocomputing*, nº71, p.733-742.
- Pulido-Calvo, I., Montesinos, P., Roldán, J. e Ruiz-Navarro, F. (2007): “Linear regressions and neural approaches to water demand forecasting in irrigation districts with telemetry systems”, *Biosystems Engineering*, nº 97, p.283 – 293.
- Richardson, H. W. (1973): *Teoría de la localización, estructuras urbanas y crecimiento regional, Economía Regional, 1a. ed. Vicens-Vives, España.*
- Rossini, P.A. (1997): “Artificial Neural Networks versus Multiple Regression in the Valuation of Residential Property”, *Australian Land Economics Review*, nº1.
- Saura, P. (1995): “Demanda de características de la vivienda en Murcia”, *Secretariado de Publicaciones de la Universidad de Murcia*, 1ª ed.
- Selim, H. (2009): “Determinants of house prices in Turkey: Hedonic regression versus artificial neural network”, *Expert Systems with Applications*, nº36, p.2843-2852.
- Tabales, J. M. N. (2007): *Mercados inmobiliarios: Modelización de los Precios, Tesis Doctoral, Departamento de Estadística, Econometría, I. O. y Organización de Empresas – Universidad de Córdoba, Espanha.*
- Thurston, J. (2002): “GIS & Artificial Neural Networks: Does your GIS Think?”, *GISVision Magazine*.

Worzala, E., M. Lenk, and A. Silva (1995): “An exploration of neural networks and its application to real estate valuation”, *The Journal of Real Estate Research*, n°2, p.185-201.

Zurada, J. M., Levitan, A. S. e Guan, J. (2006): “Non-Conventional Approaches to Property Value Assessment”, *Journal of Applied Business Research*, n°3.