

USO DE REDES NEURAS ARTIFICIAS PARA ESTIMAR VALORES VENAIS DE IMÓVEIS DA CIDADE DE SÃO LOURENÇO - MG

SACRAMENTO, Caio César Silva; ROCHA, Mailson Augusto; ÁVILA, Thiago Felipe Reis de (1); CARVALHO, Marcos Alberto de (2).

(1) Acadêmicos do oitavo período do Curso de Ciência da Computação, UNIFENAS, Alfenas.

(2) Professor do Curso de Ciência da Computação, UNIFENAS, Alfenas.

RESUMO:

O valor do Imposto Predial Territorial Urbano (IPTU) de um determinado imóvel é definido a partir do cálculo do seu valor venal. Tal valor é obtido com o uso de fórmulas e um conjunto expressivo de variáveis. Neste trabalho está sendo avaliado a eficiência das Redes Neurais Artificiais (RNA) para definição do valor venal. O conjunto de treinamento disponível possui informações de imóveis e seus respectivos valores venais da cidade de São Lourenço, MG. Inicialmente estão sendo utilizadas para o treinamento da RNA todo o conjunto de variáveis existentes e, em um segundo momento, subconjuntos serão também utilizados e avaliados. Esses subconjuntos serão obtidos com o uso dos Conjuntos Aproximados (*Rough Sets*). A *Lógica Fuzzy* está sendo também empregada no processo de normalização de variáveis numéricas substituídas por conjuntos *fuzzy*. Os resultados parciais ainda não são satisfatórios. A RNA mostrou-se capaz de calcular o valor venal aproximado, mas com uma margem de erro ainda não aceitável.

Palavras-chave: Redes Neurais, *Fuzzy*, *Rough Sets*, IPTU.

ABSTRACT:

The value of Building urban land tax (property tax) of a given property is set from the calculation of their market value. Such value is obtained with the use of formulas and an impressive set of variables. This work is being assessed the effectiveness of Artificial Neural Networks (ANN) to define the market value. The training set has property available information and their respective market values of São Lourenço, MG. Initially they are used for training the RNA entire set of existing variables and, in a second stage, sub-assemblies will also be used and evaluated. These subsets can be obtained through the use of *Rough Sets*. The *Fuzzy Logic* is also being used in numerical variáveis standardization process replaced by *fuzzy sets*. Partial results are still not satisfactory. The RNA was shown to be able to calculate the approximate market value, but with a margin of error still not acceptable.

Keywords: Neural Networks, Fuzzy, Rough Sets, Market value.

1. INTRODUÇÃO

Existem muitas formas em que a tecnologia da informação pode ser utilizada no auxílio a tomada de decisões. Quando se dispõe de uma base de dados contendo informações necessárias para uma determinada tomada de decisão, é possível, então, aplicar tecnologias que irão auxiliar, de forma automática, os usuários de tal sistema. Neste trabalho, será apresentado uma ferramenta capaz estimar os valores venais¹ de imóveis de uma cidade.

Dentre os tributos nacionais, anualmente toda prefeitura necessita fazer uma avaliação dos imóveis para cobrança de um deles, o IPTU (Imposto Predial e Territorial Urbano). Para que seja feita a avaliação, um profissional da prefeitura vai até o imóvel em questão e, de posse de algumas informações, faz a avaliação do mesmo. A partir disso, faz cálculos e chega ao valor do IPTU.

Para estimar os valores a partir das RNA, será utilizado um banco de dados fornecido pela Diretoria de Fazenda da Prefeitura Municipal de São Lourenço.

1.1 Justificativa

Para se chegar a um valor venal de um imóvel, é necessário uma série de cálculos, informações e opiniões de especialistas. Existem algumas dificuldades relacionadas à parte prática na avaliação de imóveis, como por exemplo a quantidade de informações a serem analisadas. São muitas as variáveis, aumentando assim a probabilidade de erros. Por isso, é comum a utilização de valor venal "por aproximação", onde é atribuído um valor próximo do real, porém esse valor pode ser incoerente. Se superior, prejudica o contribuinte, caso contrário, prejudica o órgão público.

A principal razão para o desenvolvimento deste software, foi criar uma ferramenta para facilitar todo esse trabalho, reduzir a margem de erro e maximizar a eficiência no acerto dos valores à serem cobrados, procurando ser conciso na avaliação, sem dar prejuízo para qualquer parte.

¹ É o preço provável que o imóvel alcançará para compra e venda à vista, diante de mercado estável e quando o comprador e vendedor têm plena consciência do potencial de uso e ocupação que o imóvel pode ser dado.

1.2 Objetivos

O objetivo geral desse trabalho é avaliar o uso das Redes Neurais Artificiais (RNA) para encontrar um padrão para tal avaliação sem o uso de fórmulas matemáticas, como é feito atualmente.

1.3 Hipótese

Após toda a análise dos dados e treinamento de uma RNA, a ferramenta irá gerar um valor venal coerente de cada imóvel, sem o uso de fórmulas matemáticas.

2 REFERENCIAL TEÓRICO

2.1 Valor venal

É definido como uma estimativa de um preço de determinados bens e tem como função servir de base de cálculo para o Poder Público na cobrança de determinados impostos.

No município de São Lourenço, o valor venal de um determinado imóvel, é calculado através de uma fórmula (FIG. 1) que determina que o valor venal (VV) é igual ao Valor do Terreno (VT) ou Valor da fração de terreno (VF) mais o Valor Predial da construção (VP)

Figura 1 – Cálculo do valor venal

$$VV = VT \text{ ou } VF + VP$$

Fonte: Código Tributário Municipal de São Lourenço, MG.

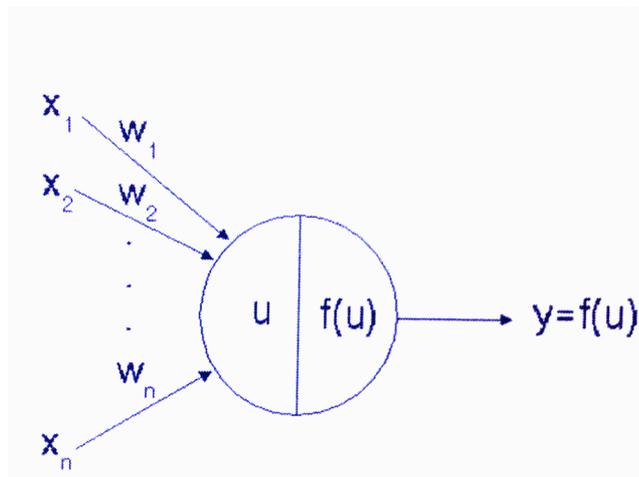
Porém, para calcular todos esses valores, é necessário uma série de informações, e para coletá-las, um funcionário da Prefeitura Municipal vai até o local do imóvel.

2.2 Redes Neurais Artificiais

Sendo inspirado pela estruturas neurais biológicas, as Redes Neurais Artificiais (RNAs) são modelos computacionais que realizam reconhecimento de padrões, por exemplo. Elas são treinadas especificamente para cada tipo de aplicação e, à medida que tem interação com o ambiente externo, vão melhorando seu desempenho.

No início da década de 40, dois estudiosos Mcculloch e Pitts (1943) desenvolveram o primeiro modelo, o MCP (FIG. 2), que simulava um neurônio artificial. No modelo, a RNA recebe variáveis de entrada x_1, x_2, x_n , onde para cada uma das entradas, há um peso correspondente w_1, w_2, w_n . A soma das entradas ponderadas pelos pesos correspondentes, geram uma saída linear u , esta sofre aplicação de uma função gerando uma saída de ativação y , que é a solução.

Figura 2 – Visão esquemática de um neurônio MCP



Fonte: REZENDE, Solange O. **Sistemas Inteligentes**, 2005. Figura 6.1.

Considerando, por exemplo, a função de ativação $f(u) = y = 0$ quando $u < \theta$ e $f(u) = y = 1$ quando $u \geq \theta$, com duas entradas binárias x_1 e x_2 com os parâmetros: $w_1 = w_2 = 1$ e $\theta = 0,5$. O neurônio ativa sua saída sempre que uma de suas entradas tenha valor igual a um, utilizando da função lógica “ou”.

O aprendizado ou treino de uma RNA é a partir de fazer e refazer os testes, ou seja, por experiência. Dessa forma, como já foi dito, ela altera os pesos dos dados de entrada para que a rede realize uma tarefa. Após treinada, a RNA ajusta

seu comportamento aprendendo a receber os dados de entrada e a dar as melhores soluções, com um desempenho cada vez melhor.

Existem algoritmos para essa alteração dos pesos e basicamente são divididos em três:

- **Aprendizado supervisionado:** onde existe um supervisor que monitora o que a rede vai receber de dado e retornar uma saída, saída essa que já deverá ser conhecida pelo supervisor. É comparado, então, o valor retornado com o valor esperado para se obter o ajuste de peso. Usado, por exemplo, na classificação de dados.
- **Aprendizado não-supervisionado:** onde não se sabe os valores de saídas desejadas e o ajuste de peso é feito com base nos valores de entrada. Usado, por exemplo, na categorização de dados.
- **Aprendizado por reforço:** onde não se sabe os valores de saída desejadas mas existe um supervisor, que verifica se a saída retornada melhorou o desempenho da rede, se sim aquela ação será executada com mais frequência, se o desempenho foi degradado, aquela ação será executada com menos frequência. Usado, por exemplo, para controle de um pêndulo invertido.

2.3 Teoria dos Conjuntos Aproximados – *Rough Sets*

Baseado no crescimento das informações e necessidade de auxílio ao tomar decisões gerenciais, foi criada uma teoria, denominada Teoria Conjuntos Aproximados (TCA), ou *Rough Sets*. Em um banco de dados, o grande volume de informações faz com que a tomada de decisão seja algo difícil sendo e que muitas dessas informações são desnecessárias para uma determinada decisão.

A TCA intitula essas informações “desnecessárias” como variáveis ou atributos irrelevantes, que por sua vez, tem como objetivo, reduzi-las, pois reduzindo a quantidade de informações, facilita a tomada de decisão. Essa redução é feita baseada na definição de redutos onde são criados subconjuntos de atributos, que fornecem o necessário para a tomada de decisão sem a necessidade de utilização de todo o banco de dados.

2.4 Lógica Fuzzy

A lógica clássica trata de valores definidos, como “ligado” ou “desligado”, “verdadeiro” ou “falso”, mas nem sempre essas variáveis booleanas irão resolver os problemas, por isso são necessários valores intermediários, como “andar alguns metros” ou “perder alguns quilos”. Para exemplificar, considerando que todos os homens acima de 1,90m sejam altos, de acordo com a lógica clássica, um homem com 1,88m não seria alto, já na lógica fuzzy ele seria “quase alto”.

Na lógica fuzzy, os valores de entrada são denominados “crisp”, depois ele sofre uma “fuzzyficação” e a saída é um valor “fuzzy”. Vários valores de entrada formam um Conjunto Fuzzy. Estes valores podem pertencer, não pertencer ou pertencer parcialmente a um conjunto Fuzzy.

Após a “fuzzyficação”, cada valor tem sua pertinência, que nada mais é do que quanto vale cada elemento no conjunto. Considerando, por exemplo, uma temperatura de 37 graus para um Conjunto Fuzzy “Frio”, “Morno” e “Quente”, ele terá as seguintes pertinências, 0.1 para frio, 0.4 para morno e 0.8 para quente.

Para que essa “fuzzyficação” aconteça, é necessário criar as Regras de Inferência, onde serão a base para o “raciocínio fuzzy”.

3 MATERIAL E MÉTODOS

Para o desenvolvimento deste trabalho, foi feito inicialmente um levantamento bibliográfico para que se obtenha informações do problema e como pode ser solucionado. Também foi necessário um levantamento de dados através da utilização de um banco de dados da cidade de São Lourenço-MG, onde consta 28.142 registros e dezenove variáveis.

Para a eliminação de variáveis supérfluas será utilizada a ferramenta ROSETTA, que é um programa baseado em TCA e cobre todas as etapas do processo com opções para discretização, cálculo de redutos e geração de regras.

Foi utilizado o ambiente Microsoft Visual Studio 2012, com a linguagem de programação C#, para criação do software e treinamento da RNA.

Do total de registros, o usuário tem a opção de escolher o tamanho do conjunto de treinamento da RNA e outros vários para execução de testes, onde todos foram selecionados por um sorteio dirigido pela ordem da área.

A lógica difusa foi usada para melhorar o entendimento da RNA, sendo utilizada na variável área. Além disso, foi necessário a normalização dos dados, que consiste em entregar para a RNA valores entre 0 e 1, para tal foi utilizado da fórmula

Automatizaremos a parte de testes.
Após os testes, os resultados serão avaliados e apresentados.

4 DESENVOLVIMENTO

4.1 Características

Para automatizar e facilitar o cálculo venal na cidade de São Lourenço-MG, foi criado o ELVVIS.

Com o software, o usuário tem a possibilidade de visualizar o banco de dados, criar um Conjunto de Treinamento, treinar uma nova ou carregar um treinamento de uma RNA, e entrar com novos dados para ser estimado o valor venal.

4.2 Complicações e soluções

Sem a discretização do banco, não é possível o treinamento da RNA, portanto iniciou-se o processo de discretização dos dados. Durante o desenvolvimento e os testes com o primeiro conjunto de treinamento, foi constatado que os valores que a RNA estava retornando, não eram compatíveis com os valores reais. A partir daí foi iniciado um processo de facilitação, para a RNA entender melhor o problema proposto.

Visando o melhor entendimento da RNA, o banco de dados foi alterado. Foi criado uma nova tabela onde as variáveis com mais de um tipo foram divididas e realocadas, onde cada tipo virou uma nova variável booleana.

- A Utilização se dividiu: uResidencia, uComercio, uPrestacaoServico, uEscola, uCulto, uPostoServico, uHotel, uHospital, uServicoPublico, uSupermercado, uAtividadeC, uIndustria, uSemUtilizacao.
- A Zona Fiscal se dividiu: zf1, zf2, zf3, zf4, zf5, zf6, zf7, zf8, zf9, zf10, zf11, zf12, zf13, zf14, zf15, zf16, zf17, zf18, zf19.
- O posicionamento na quadra se dividiu: posQdUmaF, posQdDuasF,

posQdDuasFEsquina, posQdMaisDuasF. A Pedologia se dividiu: pedologiaFS, pedologiaRAI.

- A Topografia se dividiu: topografiaPlano, topografiaAL, topografiaAM, topografiaAA, topografiaDL, topografiaDM, topografiaDA.
- O Acabamento se dividiu: acabamentoPrecario, acabamentoPopular, acabamentoMedio, acabamentoFino, acabamentoLuxo.
- A Conservação se dividiu: conservacaoBoa, conservacaoMedia, conservacaoRuim, conservacaoPessima, conservacaoEmR.
- A Situação se dividiu: situacaoFrente, situacaoFundo, situacaoSubsolo, situacaoGaleriaComVisao, situacaoGaleriaSemVisao.

Com enormes discrepâncias entre valores, a RNA estava se “estressando” e não estava aprendendo. Surgiu então a idéia de usar a Lógica Fuzzy, criando um Conjunto Fuzzy para a variável Área, que foi dividida entre: área muito pequena, pequena, média, grande e muito grande. Assim, os valores crisp foram “fuzzyficados”.

A partir disso é necessário criar o Conjunto de Treinamento, onde o usuário deve escolher qual a porcentagem de dados será utilizada, pois já foi constatado que um número muito exagerado de dados atrapalha o aprendizado da RNA.

Com o Conjunto de Treinamento criado, será aplicada uma bateria de testes.

4.3 O Sistema

O software, oferece ao usuário opções para “Visualizar Banco de Dados”, “Criar Conjunto de Treinamento”, “Visualizar Conjunto de Treinamento”, “Excluir Conjunto de Treinamento”, “Treinar a RNA” , “Testar Treinamento” e “Estimar Valor Venal”.

O usuário tem a possibilidade de visualizar o banco de dados já carregado pelo sistema, como mostra na FIG. 3.

Figura 3 – Tela de visualização do Banco de Dados

	Area	Unidades	Fracao	Testada	Utilizacao	Calçamento	Iluminacao	Lim
▶	98,43	1	0,00	19,85	RESIDENCIA	S	S	S
	184,83	2	90,07	9,15	COMERCIO	S	S	S
	184,83	2	94,76	9,15	RESIDENCIA	S	S	S
	181,57	1	0,00	9,00	RESIDENCIA	S	S	S
	230,00	1	0,00	11,50	RESIDENCIA	S	S	S
	408,55	1	0,00	20,00	RESIDENCIA	S	S	S
	1.200,00	1	0,00	30,00	RESIDENCIA	S	S	S
	377,00	3	181,32	47,20	RESIDENCIA	S	S	S
	377,00	3	91,49	47,20	RESIDENCIA	S	S	S
	377,00	3	104,19	47,20	TERRENO VAG...	S	S	S
	270,00	1	0,00	7,00	RESIDENCIA	N	S	S
	130,00	1	0,00	3,00	TERRENO VAGO	N	S	S
	170,00	1	0,00	3,05	RESIDENCIA	N	S	S
	159,85	1	0,00	6,95	RESIDENCIA	S	S	S
	400,00	2	143,69	10,00	RESIDENCIA	N	S	S
	400,00	2	256,31	10,00	RESIDENCIA	N	S	S
	126,75	1	0,00	6,50	RESIDENCIA	S	S	S
	999,00	2	565,40	22,00	RESIDENCIA	S	S	S
	999,00	2	433,60	22,00	RESIDENCIA	S	S	S
	100,00	1	0,00	10,00	RESIDENCIA	S	S	S
	214,67	1	0,00	44,15	RESIDENCIA	S	S	S
	400,00	1	0,00	10,00	RESIDENCIA	S	S	S
	422,01	1	0,00	10,10	RESIDENCIA	S	S	S

Na FIG. 4, o usuário treina a RNA a partir do Conjunto de Treinamento que foi criado e salva o treinamento.

Figura 4 – Tela de Treinamento da RNA de imóveis não construídos

	Area	Unidades	Fracao	Testada	Utilizacao	Calcamento	Iluminacao	Limp
*								

Nº máximo de iterações: 7000 Taxa de Aprendizagem: 0.8

Neurônios Camada de Entrada: 8 Taxa de Inércia: 0.2

Neurônios Camada Oculta: 8 Erro: 0.0100

Treinar

O usuário tem, também, a possibilidade de testar o treinamento da RNA, onde verifica-se a qualidade do treinamento. Se não houve um resultado satisfatório, a RNA pode ser treinada novamente.

Finalmente, na FIG. 5, o usuário entra com os dados e tem como saída o valor venal.

Figura 5 – Tela para estimar o valor venal

5 RESULTADOS E DISCUSSÃO

Com o banco de dados “bruto” e com um alto número de registros, o sistema não foi eficiente, devolvendo valores que não condizem com a realidade, porém os resultados negativos já eram esperados.

Assim, estão sendo investigadas alternativas para a resolução dos problemas.

6 CONCLUSÃO

Com o software em sua versão inicial, ainda não foi possível estimar valores venais de forma precisa, mas com todas as mudanças efetuadas, espera-se uma melhora significativa na saída da RNA.

REFERÊNCIAS

CARVALHO, Marcos Alberto de. **Discretização de atributos contínuos em sistemas de informação utilizando algoritmos genéticos para a aplicação da teoria dos conjuntos aproximados.** Dissertação (Mestrado em Ciência e Tecnologia da Computação) – UNIVERSIDADE FEDERAL DE ITAJUBÁ, Itajubá – MG, 2010.

CÓDIGO TRIBUTÁRIO MUNICIPAL DE SÃO LOURENÇO – MG. **Lei complementar 001/2010.** Anexo I.

REZENDE, Solange Oliveira (Coord.). **Sistemas inteligentes: Fundamentos e Aplicações.** Barueri, SP: Manole, 2005.

THOFEHRN, Ragnar. **Avaliação em massa de imóveis urbanos: Para cálculo de IPTU e ITBI.** [S.l.]: PINI, 2010.