Construção de modelos do mercado imobiliário para análise de viabilidade com regressão e sistemas de regras difusas

Construction of real estate models for feasibility analysis using regression and fuzzy rules

Marco Aurélio Stumpf González
Carlos Torres Formoso

Resumo

O estudo de viabilidade para o lançamento de um novo empreendimento muitas vezes é realizado com base em critérios simplificados ou subjetivos. A análise cuidadosa do mercado é uma atividade essencial nesse processo, sendo importante para reduzir parte do risco. Podem-se obter estimativas mais precisas das receitas mediante a construção de modelos hedônicos com dados coletados no mercado. Entretanto, a avaliação do valor de mercado dos imóveis por esse processo também enfrenta dificuldades, tais como imprecisões nas transições entre diferentes submercados. As regras difusas constituem uma forma de consideração dessas imprecisões e permitem incrementar a precisão dos modelos de avaliação. Neste trabalho apresenta-se uma aplicação de sistemas baseados na técnica de regras difusas com ajustamento através de algoritmos genéticos. Foram desenvolvidos dois sistemas de regras difusas, sendo o primeiro baseado na área total dos imóveis e o segundo baseado na sua localização. Esses modelos foram comparados com o tradicional modelo hedônico de regressão. Os resultados indicam que essa técnica pode contribuir para melhorar a qualidade das avaliações no mercado de imóveis.

Palavras-chave: Análise de viabilidade. Regras difusas. Avaliação em massa. Mercado imobiliário. Algoritmos genéticos. Segmentação de mercado.

Abstract

The feasibility study for launching a new property development is often carried out using simplified or subjective techniques. Careful market analysis should be an essential part of this process, in order to reduce part of the risk involved. More precise price estimations may be obtained using hedonic price models, based on sales market data. However, estimating the market value of real estate through this process has also some drawbacks, such as imprecision of the boundaries between different sub-markets. Fuzzy rules may be used for dealing with this type of imprecision, making it possible to improve the precision of market appraisal models. This paper presents an application of fuzzy rule-based systems, which have been adjusted by genetic algorithms. Two fuzzy systems were developed. The first one was based on the property area, and the second one was based on its location. Both were compared to a traditional hedonic regression model. The results have indicated that fuzzy rules may improve the quality of real state market analysis.

***Keywords**: Feasibility analysis. Fuzzy rules. Mass appraisal. Property market. Genetic algorithms. Market segmentation.*

Marco Aurélio Stumpf González
Faculdades de Engenharia Civil e de Arquitetura e Urbanismo da Universidade do Vale do Rio dos Sinos
Av. Unisinos, 950
São Leopoldo – RS – Brasil
CEP 93022-000
Tel.: (51) 3591-1122, ramal 1691
Fax.: (51) 3590-8172
E-mail: mgonzalez@unisinos.br

Carlos Torres Formoso
Núcleo Orientado para Inovação da Edificação
Universidade Federal do Rio Grande do Sul
Av. Osvaldo Aranha, 99, 3º andar, Centro
Porto Alegre – RS - Brasil
CEP 90035-190
Tel.: (51) 3316-4084
Fax.: (51) 3316-4054
E-mail: formoso@.ufrgs.br

Recebido em 24/05/05
Aceito em 08/11/06

Introdução

O mercado imobiliário representa um dos mais importantes segmentos da economia nacional (HARVEY, 1996). As características especiais do mercado imobiliário tornam difícil o processo de decisão de investimento ou de lançamento de novas construções, principalmente na etapa de análise de viabilidade. Muitas vezes, a decisão é tomada pelo empreendedor de forma subjetiva, de acordo com sua experiência e percepção das condições momentâneas do mercado, sem ter como base uma análise mais criteriosa, fundamentada em dados. Como lembra Lavender (1990), o agente busca atingir algum benefício (financeiro, social ou de outro tipo) com o empreendimento e antes de decidir deve desenvolver uma avaliação cuidadosa para assegurar que o projeto proposto pode efetivamente atingir seus objetivos. No caso da construção civil, quando a oportunidade de uma nova construção é investigada, denomina-se essa avaliação de "análise de viabilidade".

Podem ser estimados modelos numéricos do mercado com base na teoria de preços hedônicos, através do uso da análise de regressão (ROBINSON, 1979; SHEPPARD, 1999). Esses modelos têm como objetivo a estimação do valor de mercado dos imóveis, utilizando uma amostra de dados de transações do segmento de mercado de interesse, mas contam com uma parcela de erro. Parte desses erros deve-se à imprecisão do mercado imobiliário, e outra parte aos próprios modelos do mercado, geralmente construídos mediante análise de regressão. As regras difusas constituem uma forma interessante de consideração das imprecisões típicas do mercado imobiliário, tais como as transições entre regiões ou tipos de imóveis contíguos (interfaces entre submercados). Neste trabalho apresentam-se os sistemas de regras difusas como uma alternativa aos modelos de regressão tradicionais, investigando a extração de regras difusas a partir de uma base de dados, com ajustamento das regras através de algoritmos genéticos, formando um Sistema Baseado em Regras Difusas Evolucionário (SBRDE). Este trabalho tem ainda o objetivo de demonstrar a utilidade dos sistemas baseados em regras difusas na área da construção, como uma ferramenta para análises numéricas e para a solução de problemas. Embora no caso apresentado os resultados sejam similares, é importante considerar os algoritmos genéticos como uma técnica útil para a solução de problemas com outras características. O trabalho organiza-se como segue: a próxima seção apresenta os sistemas de regras difusas. Em seguida, apresentam-se a aplicação realizada e os resultados obtidos, seguidos pelas conclusões.

Sistemas baseados em regras difusas

Os sistemas baseados em regras difusas (SBRD) apresentaram um grande desenvolvimento teórico e em termos de aplicações nos últimos anos, especialmente com sistemas híbridos. Os SBRD consistem em uma base de regras que utilizam a lógica difusa nas partes precedentes ou antecedentes das regras, e uma das características interessantes desses sistemas é a conjugação de efeitos de várias regras para a obtenção do resultado final (CORDÓN et al., 2001). Esses sistemas são uma extensão dos sistemas baseados em regras clássicas, usando as regras difusas de forma a habilitá-los para aplicações em áreas que apresentam incerteza ou imprecisão (CORDÓN; HERRERA, 1999; ALCALÁ et al., 2000; CORDÓN et al., 2001).

Um conjunto difuso é um tipo de conjunto no qual existe uma progressão gradual de pertinência para não-pertinência ou, mais precisamente, no qual um objeto pode ter um grau de pertinência intermediário entre a unidade (pertinência total) e zero (não-pertinência). Neste caso, a função característica dos conjuntos clássicos ($\varphi_A(x)=\{0, 1\}$) é substituída por uma relação de pertinência ($\mu_A(x)=[0, 1]$), a qual considera também os valores intermediários, ou seja, inclui os casos em que o objeto pertence parcialmente a dois (ou mais) conjuntos (ZADEH, 1965; NGUYEN; WALKER, 2000). Os valores de pertinência expressam os graus nos quais cada objeto é compatível com as propriedades que caracterizam o conjunto em questão (características distintivas da coleção), ou seja, esses valores expressam o grau de associação do elemento com a característica representada pelo conjunto. Assim, um conjunto difuso é caracterizado por uma função de pertinência $\mu_A(x)$, que assume um valor qualquer no intervalo [0,1]. Fica claro que um conjunto difuso é uma generalização do conceito clássico de conjuntos, no qual a função característica assumia apenas dois valores, {0, 1}. Assim, se A é um conjunto difuso, a relação pode ser expressa como na Equação 1 (KACPRZYK, 1997; CIOS et al., 1998).

$\mu_A(x) = 1$, se x $\in$ (totalmente) A (1)
$\mu_A(x)= \mu_x$, se x pertence parcialmente a A, com $0<\mu_x<1$
$\mu_A(x)= 0$, se x $\notin$ (totalmente) A

Por exemplo, se existem dois conjuntos difusos, denominados "branco" (B) e "preto" (P), e são apresentados elementos intermediários, com quantidade variável de branco e de preto (ou seja, tons de cinza), algumas das instâncias desses

conjuntos difusos podem ser exemplificadas como na Figura 1.

Um elemento tal como x=6 terá pertinência parcial de $\mu_p(6)=0,25$ e $\mu_B(6)=0,75$, o que significa que está mais relacionado com B do que com A. A teoria clássica de conjuntos não permite uma resposta adequada a essa questão.

Um SBRD tem dois componentes: (a) o sistema de inferência, que realiza o processo de cálculo necessário para gerar uma saída quando determinada entrada é especificada; e (b) a base de conhecimento, que representa o conhecimento sobre o problema a ser resolvido, constituída por uma coleção de regras, as quais geralmente são obtidas a partir de uma coleção de casos (exemplos) do domínio. Existem vários algoritmos de aprendizagem, tais como redes neurais, algoritmos genéticos e clusterização (ALCALÁ et al., 2000). Uma regra difusa tem o seguinte formato geral (Equação 2):

R_i: SE X é A_i ENTÃO Y é B (2)

Onde: a parte "SE X é A_i" é chamada de antecedente, e a parte "ENTÃO Y é B" é chamada de conseqüente da regra *i*, A e B são conjuntos difusos, X é a entrada apresentada ao sistema e Y é o resultado calculado. Um sistema de regras é composto de um conjunto de regras de formato semelhante à Equação 2. Existem basicamente dois tipos de SBRD: Mamdani e TSK. O primeiro considera variáveis lingüísticas na parte conseqüente, enquanto o tipo TSK utiliza como resultado uma função numérica das entradas (ALCALÁ et al., 2000).

O modelo TSK é composto de antecedentes representados por variáveis lingüísticas, sendo os conseqüentes funções das variáveis de entrada, geralmente representados por funções lineares (CORDÓN et al., 2001), tal como as regras das Equações 3 e 4.

R_i: SE X_1 é A_1 e ... e X_n é A_n ENTÃO $Y_i=f(X_1, ..., X_n)$ (3)

R_i: SE X_1 é A_1 e ... e X_n é A_n ENTÃO $Y_i=p_0 + p_1X_1 + ... + p_nX_n$ (4)

Onde: X_j são as variáveis de entrada, A_j são conjuntos difusos especificando seu significado, Y_i é o resultado da regra *i*, e os parâmetros p_k são números reais (CORDÓN; HERRERA, 1999; CORDÓN et al., 2001). Os A_j podem ser variáveis lingüísticas associadas com conjuntos difusos ou variáveis difusas, diretamente. O resultado, considerando-se uma base de conhecimento e considerando-se um conjunto de regras, é obtido por uma média ponderada das regras do sistema (ALCALÁ et al., 2000; CORDÓN et al., 2001).

As maiores vantagens dos sistemas TSK são que consistem de um conjunto compacto de equações e que os parâmetros p_k podem ser estimados por qualquer processo, tais como regressão múltipla, algoritmos genéticos ou mesmo pela indicação de especialistas. A estrutura básica do sistema do tipo TSK está representada na Figura 2 (CORDÓN et al., 2001):

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\mu_P(x)$	0,00	0,05	0,10	0,15	0,20	0,25	0,30	0,40	0,50	0,60	0,70	0,75	0,80	0,90	1,00
$\mu_B(x)$	1,00	0,95	0,90	0,85	0,80	0,75	0,70	0,60	0,50	0,40	0,30	0,25	0,20	0,10	0,00

Figura 1 – **Exemplo de conjuntos difusos**

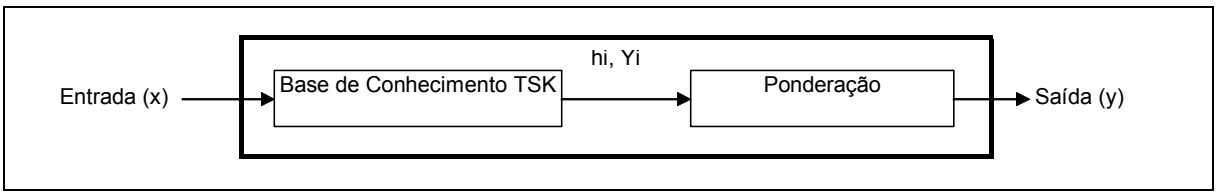

Figura 2 – **Estrutura de um sistema de regras do tipo TSK (adaptado de Cordón et al., 2001)**

Os sistemas de regras TSK são baseados em informação numérica e são especialmente úteis quando não existe ou quando existe pouco conhecimento disponível. De qualquer forma, o conhecimento especializado é útil para definir os rótulos das variáveis lingüísticas (*labels*), especificar seu significado ou identificar os possíveis estados do sistema (combinações de entradas e saídas). Os conseqüentes nas regras TSK (as equações lineares) geralmente são gerados por meio de um processo semi-automatizado.

Nem sempre esse formato é adequado para representar conhecimento introduzido diretamente pelo especialista. Porém, o conhecimento pode ser introduzido por meio de uma pequena modificação: quando uma regra é fornecida pelo especialista, com o conseqüente na forma "Y é B", pode-se entender como "$Y=p_0$". Este tipo de regra é conhecido como "regra TSK simplificada" ou "regra TSK de ordem zero" (ALCALÁ et al., 2000).

Desenvolvimento de um SBRD

Basicamente, o desenvolvimento do sistema consiste em encontrar um conjunto de regras difusas que sejam aptas a reproduzir o comportamento de entrada e saída do sistema, com base apenas em um conjunto de dados, composto de vetores de entrada e saída. A modelagem é realizada geralmente em duas etapas. A primeira consiste na escolha das variáveis relevantes e geração de uma primeira aproximação para o modelo difuso que descreve o sistema, incluindo a determinação do número de regras necessárias. O resultado dessa etapa é uma coleção de regras que pode ser vista como uma estimativa da base final de regras. O segundo passo consiste em sintonizar o conjunto inicial de regras, obtendo-se a base final (GÓMEZ-SKARMETA; JIMÉNEZ, 1997; CORDÓN et al., 2001).

Todavia, os sistemas de regras difusas não são capazes de aprender sozinhos as regras. Em geral, os sistemas difusos são construídos através de abordagens híbridas, usando-se algoritmos genéticos ou redes neurais na etapa de aprendizagem das regras. Existem duas abordagens básicas para o treinamento. Na abordagem Pittsburgh, todas as regras do sistema são extraídas dos dados simultaneamente. No formato Michigan, cada regra é obtida individualmente, construindo-se o sistema progressivamente. Em geral, a abordagem Pittsburgh proporciona sistemas mais consistentes. As vantagens do segundo formato são a flexibilidade do sistema (atualizado mais facilmente) e o menor esforço de computação, visto que é treinada uma regra por vez (CORDÓN et al., 2001).

Há várias formas de se obter a base de regras, geralmente contando com o apoio de outras técnicas, tal como os algoritmos genéticos (AG), os quais são utilizados neste estudo. Os AG são procedimentos de busca ou otimização inspirados na regra natural dos sistemas biológicos de "sobrevivência dos mais aptos". Eles empregam uma busca aleatória dirigida, buscando uma solução ótima global. Já foi provado que os AG podem encontrar soluções quase-ótimas para problemas complexos. Os AG geram aleatoriamente uma população inicial de cromossomos, as quais são soluções em potencial para o problema. Os cromossomos são codificados como números reais ou cadeias binárias. O algoritmo faz evoluírem essas soluções em sucessivas gerações através de mecanismos de seleção e reprodução, até que seja atingido o nível de erro desejado ou o número máximo de ciclos especificado. Os indivíduos mais aptos podem ser selecionados diretamente (seleção elitista). As gerações seguintes também são definidas por meio de cruzamento (misturando partes de dois cromossomos) ou mutação (uma parte do cromossomo é alterada aleatoriamente). A cada nova geração, os indivíduos (cromossomos) são testados com respeito a uma função de aptidão ou ajustamento (*fitness function*), e os cromossomos com maior nível de aptidão têm maior probabilidade de sobrevivência (CORDÓN et al., 2001; GOLDBERG, 1989).

Aplicações de sistemas de lógica difusa

A lógica difusa tem emprego freqüente em sistemas de controle e em sistemas de regras difusas. Geralmente é aplicada em conjunto com outras técnicas, tais como redes neurais, raciocínio baseado em caso e algoritmos genéticos, em sistemas híbridos, para proporcionar a aprendizagem da base de regras ou sintonizar o sistema (BONISSONE et al., 1999; CORDÓN et al., 2001). Bonissone et al. (1998, 1999) descrevem uma aplicação no mercado imobiliário. O sistema de avaliações apresentado por esses autores tem dois componentes. No primeiro, a lógica difusa auxilia na definição da arquitetura de redes neurais. No segundo componente, ela é utilizada na identificação de casos similares para um subsistema de raciocínio baseado em casos.

Byrne (1995) propôs a utilização de lógica difusa para levar em conta os elementos de risco e incerteza presentes nas avaliações. Bagnoli e Smith (1998) empregaram lógica difusa para considerar as imprecisões constantes nas medidas

dos imóveis, através da "fuzzificação" das variáveis decorrentes de julgamentos dos avaliadores, numa tentativa de reproduzir o mecanismo de decisão dos agentes, que contam com informação imprecisa.

Na construção civil também existem algumas aplicações. Por exemplo, Leu et al. (1999) desenvolveram um sistema para planejamento de obras utilizando lógica difusa para considerar as incertezas na duração das atividades. A otimização do planejamento em função das restrições de recursos foi realizada com algoritmos genéticos. Karray et al. (2000) utilizaram lógica difusa para o planejamento de canteiro de obras, também em um sistema híbrido com algoritmos genéticos. Soh e Yang (1996) e Yang e Soh (2000) apresentaram dois sistemas para cálculo de estruturas metálicas: o primeiro utiliza algoritmos genéticos e o segundo aplica programação genética, ambos incorporando conhecimento de especialistas no processo de otimização das estruturas.

Em uma abordagem diferente, Chao e Skibniewski (1998) descreveram um sistema para avaliação de alternativas tecnológicas para a construção baseado em lógica difusa. O exemplo apresentado é sobre métodos alternativos para operação de formas em prédios altos. A motivação para o estudo é que as técnicas formais de decisão, tais como as baseadas na teoria da utilidade, podem ser substituídas com vantagem por um sistema com lógica difusa. O trabalho mostra uma abordagem de avaliação de novas tecnologias de construção, de forma a embasar decisões consistentes.

Aplicação

Para demonstrar a montagem dos modelos de mercado, foi elaborado um estudo com dados de vendas de apartamentos em Porto Alegre. Três modelos foram construídos e comparados. O primeiro é um modelo hedônico baseado em regressão, no formato tradicional, e os outros dois são sistemas baseados em regras difusas. A regressão foi utilizada como referência em função de ser a técnica mais utilizada atualmente em avaliação coletiva de imóveis, sendo bem conhecida e razoavelmente mais simples, inclusive em termos de disponibilidade de software.

Foram estimados dois tipos de conjuntos difusos, utilizando-se em ambos algoritmos genéticos para extrair as regras do banco de dados, o que gerou sistemas baseados em regras difusas evolucionários (SBRDE). Dois tipos de sistemas de regras difusas foram testados. No primeiro caso foi estimado um SBRDE com funções de pertinência baseadas na área total dos imóveis. Foram investigados sistemas de regras com 3 a 7 regras, determinados através da abordagem Pittsburgh.

No segundo caso, foi construído um sistema de regras baseado em funções de pertinência que contemplam a localização. A fundamentação é a continuidade espacial do mercado, considerada mediante a união difusa das várias regras, cada uma delas "especializada" em determinada região, mas também contribuindo para a estimação nas demais, principalmente nas regiões contíguas. Esse formato utilizou a abordagem Michigan, que permite maior flexibilidade na definição do número de regras, as quais são estimadas individualmente.

Nas duas abordagens, uma das vantagens observadas é a possibilidade de utilizar-se o conhecimento disponível, tais como modelos determinados por outras técnicas, como estimativa inicial (semente) para a geração das populações de teste. Outra característica é a flexibilidade do sistema. Os dois tipos de sistemas difusos utilizaram regras TSK, as quais são formadas por equações lineares e podem ser vistas como modelos hedônicos.

A base de dados consiste de dados extraídos de declarações do imposto de transmissão (ITBI) sobre vendas de 31.277 apartamentos, vendidos entre agosto de 1998 e agosto de 2001. Os dados disponíveis incluem o preço estimado pela prefeitura (Preço), as áreas total (Artocons) e privativa (Arcopriv), o padrão de qualidade da construção (usando quatro variáveis binárias – Pop, Med, Fin e Lux), a idade, a data da venda e o endereço. Algumas características importantes, como número de dormitórios e existência de vagas de estacionamento, não estavam disponíveis e, tendo em vista o tamanho da amostra, não foi possível coletar tais informações. Detalhes sobre os dados e procedimentos adotados podem ser obtidos em González (2002) e Soibelman e González (2002).

A variável de localização (Bairro) foi considerada pela identificação dos imóveis aos bairros da cidade, utilizando uma medida definida pelos autores baseada em renda familiar, qualidade média da habitação e qualidade de vizinhança em cada bairro. Em Porto Alegre, os limites dos bairros são definidos por lei municipal. Também foi considerada a posição espacial dos imóveis através de suas coordenadas, com origem na sede da prefeitura municipal. Com essa referência, foram calculadas as distâncias aos principais centros comerciais (shoppings e outros pólos tradicionais), bem como a distância ao centro da cidade (representando o *Central Business District* - CBD), ambas medidas em quilômetros. Nos

modelos determinados, a variável binária Pop não foi incluída, mantendo-se as demais variáveis binárias relacionadas com o padrão de construção do imóvel (Med, Fin e Lux). O mês da venda foi incluído considerando-se uma escala contínua, iniciando no mês da venda mais antiga (Mês=1 para agosto de 1998). Por fim, foi calculada a razão entre a área privativa e a área total (Razão), e foi incluída uma variável binária para considerar a existência de vista panorâmica (Vista). Os preços foram corrigidos monetariamente para a data de agosto de 2001 usando-se o IGP-DI.

Os dados foram pré-processados, e 56 casos foram removidos em função de erros, seguindo a metodologia apresentada em González (2002) e Soibelman e González (2002). A presença de casos não relevantes para a análise foi investigada utilizando-se a abordagem envelope (*wrapper approach*), usando a análise de regressão em modelos preliminares. Este método foi originalmente proposto por Kohavi e John (1997) para a seleção de atributos e foi utilizado para a seleção de casos como uma extensão, por analogia. A seleção foi realizada utilizando-se os erros obtidos em modelos iniciais. Casos com grandes resíduos foram removidos, considerando-se um limite de três desvios padrão (limite adotado em função do tamanho e da variedade da amostra). O procedimento de Box-Cox foi empregado para investigar transformações numéricas nas variáveis, resultando na aplicação de um expoente 3/4 na variável dependente. Ao final desse processo, os 26.579 casos resultantes foram divididos em dois arquivos, para treinamento e teste dos modelos, respectivamente contendo 80% e 20% dos dados. Essa é uma prática comum em estudos com inteligência artificial e permite a validação dos modelos (HAYKIN, 2001). As características da amostra estão apresentadas na Tabela 1. Variáveis identificadas com a letra "C" têm variação contínua, enquanto aquelas identificadas com "B" são variáveis binárias.

Regressão

Foram utilizados os limites de significância recomendados pela norma brasileira de avaliação de imóveis, NBR-14653-2 (ABNT, 2004), que pode ser utilizada como referência para estudos sobre o mercado imobiliário. Para o nível de qualidade superior previsto nessa norma, os limites consistem em 10% de significância para as variáveis (teste de hipótese usando a estatística *t*, no teste bicaudal) e 1% para o modelo (análise de variância com a distribuição *F*), além da análise dos pressupostos e demais condições dos modelos. Os resultados estão apresentados na Equação 5, a seguir.

$$\text{Preço} = (515{,}724 + 35{,}260*\text{Artocons} + 883{,}674*\text{Razão} + 510{,}359*\text{Med} + 1.734{,}842*\text{Fin} + 4.251{,}930*\text{Lux} - 27{,}447*\text{Idade} + 20{,}615*\text{Bairro} - 43{,}290*\text{Centro} - 90{,}022*\text{Comércio} + 1.618{,}714*\text{Vista} - 19{,}864*\text{Mês})^{4/3} \quad (5)$$

Os sinais dos coeficientes são os esperados (conforme conhecimento anterior sobre o mercado em análise), para as variáveis incluídas. Todos os atributos apresentados na Equação 6 tiveram significância muito inferior a 10%. No caso do tempo (Mês), o coeficiente negativo indica a progressiva diminuição dos preços no período coberto pela amostra. O ajustamento do modelo, com coeficiente de determinação (ajustado para os graus de liberdade) de 0,95, indica que 95% das variações da variável dependente podem ser explicadas pelo modelo apresentado. O exame dos pressupostos revelou que o modelo atende aos requisitos de normalidade e de homocedasticidade, bem como aos demais pressupostos e condições da regressão (GUJARATI, 2000; GONZÁLEZ, 2003). O valor calculado para o teste de variância do modelo (F_{calc}=41.969) também foi expressivamente superior ao mínimo exigido pela norma. Também foram examinados os resultados gerais do modelo, através dos erros percentuais (EA) e dos erros padronizados (EP), conforme dados apresentados na Tabela 2.

Sistemas de regras difusas

Sistema difuso com base na área total

Este sistema tem regras no formato da Equação 6. A parte antecedente das regras apresenta conjuntos difusos em função da área, e a parte conseqüente tem equações do tipo TSK. O conjunto de regras Ri é estimado simultaneamente através de algoritmos genéticos (abordagem Pittsburgh):

$$\text{SE Área total é } A_i \text{ ENTÃO } \text{Valor}_i = (\text{modelo}_i) \quad (6)$$

As funções de pertinência Ai foram definidas em formato triangular, para as funções centrais, com as funções dos extremos em formato trapezoidal. No caso mais geral, com 7 parcelas, o conjunto de funções de pertinência assume um formato como o da Figura 3, na qual os imóveis estão divididos em "muito pequenos" (mP), "pequenos" (P), "pequeno-médios" (PM), "médios" (M), "médio-grandes" (MG), "grandes" (G) e "muito grandes" (mG). Variando as áreas-limite, pode-se obter diferentes configurações de sistemas, com diferentes coberturas para cada regra (faixas de atuação das regras).

Variável	Tipo	Unidade	Min	Max	Média	σ
Preço	C	R$	2271,61	1603531,00	74310,64	86765,11
Artocons	C	m^2	10,59	763,36	79,92	49,07
Razão	C	–	0,12	1	0,79	0,11
Pop	B	–	0	1	0,16	0,37
Med	B	–	0	1	0,76	0,43
Fin	B	–	0	1	0,08	0,27
Lux	B	–	0	1	$1,1*10^{-3}$	$3,2*10^{-2}$
Idade	C	anos	0	81	22,30	11,59
Bairro	C	–	13	88	29,78	13,67
Centro	C	km	0,02	16,28	5,45	3,68
Comércio	C	km	0,09	10,64	2,31	1,35
Vista	B	–	0	1	$8,3*10^{-4}$	$2,8*10^{-2}$
Mês	C	–	1	37	21,97	9,33
Coordenada X	C	km	–2,95	13,50	3,85	3,86
Coordenada Y	C	km	–14,00	5,74	-1,66	3,49

Tabela 1 – **Características dos dados utilizados**

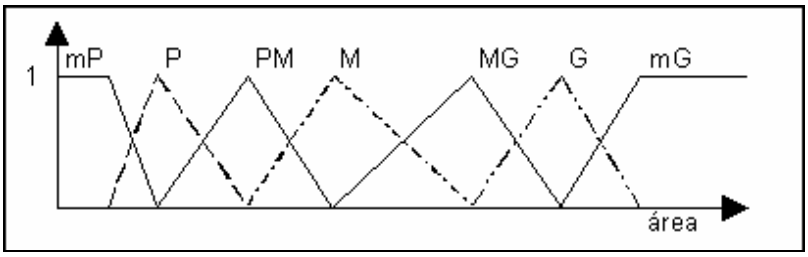

Figura 3 – **Exemplo de conjuntos difusos para a área total**

a_1	...	a_{12}	...	g_1	...	g_{12}	h_1	...	h_7

Figura 4 – **Aspecto dos indivíduos utilizados nos algoritmos genéticos para regras difusas**

Dados	EA	EP	Casos
Treino	18,45%	18.957	24.289
Teste	19,14%	19.854	6.074

Tabela 2 – **Resultados obtidos com o modelo de regressão múltipla**

Os limites de cada um dos conjuntos difusos (áreas mínima e máxima) foram determinados simultaneamente com os coeficientes dos modelos (codificados no mesmo cromossomo), porém baseados apenas em seleção (os valores originais, definidos aleatoriamente, não foram alterados). Essa estratégia foi preferida em função da maior estabilidade de evolução dos sistemas, detectada em testes iniciais. Os indivíduos foram definidos com a seguinte estrutura (Figura 4), onde os conjuntos a1,...,a12 a g1,...,g12 representam os coeficientes para a parte conseqüente de cada regra (cada um com 11 coeficientes para as variáveis e um intercepto), e h1,...,h7 são os limites das funções de pertinência (em termos de área total), conforme a Figura 3. Cada indivíduo representa um sistema de regras completo. Por exemplo, os cromossomos para os conjuntos de três regras têm 39 genes (3x12+3), enquanto os de sete regras têm 91 genes (7x12+7), seguindo o formato da Figura 4.

A geração da população inicial utilizou o conhecimento disponível. As regras iniciais foram baseadas em equações de regressão desenvolvidas com base em 3, 5 e 7 subamostras, divididas por clusterização com base no critério "área total" (gerando os limites h_1, ..., h_7). Em todos os casos

foram aplicadas variações aleatórias sobre os coeficientes para a geração da população inicial (dentro da faixa ± 50%).

Em função de as regras envolverem diferentes tamanhos de imóveis, com diferentes valores totais, a competição através do erro padrão (EP) privilegia os imóveis menores, os quais têm erros menores, em termos absolutos, e também são a maioria em número de casos. Assim, a evolução das regras tende a produzir um conjunto bem adaptado aos imóveis pequenos mas progressivamente menos adequado para os médios e grandes. Para evitar essa situação, a função de ajustamento (*fitness function*) baseou-se no erro absoluto percentual médio (EA) e teve o seguinte formato (Equação 7):

$$Fi=1/(1+EAi), \quad (7)$$

com $EAi=\Sigma i,j\ [|Yj-Yhi,j| / Yj * 100\ ]$,

Onde: Fi é a estimativa de fitness para a regra i, EAi é o erro absoluto percentual médio calculado para a regra i com os dados de treinamento, Yj é o preço observado, e Yhi,j é o preço estimado pela regra i para o caso j.

Os operadores genéticos utilizados foram cruzamento e mutação. Os parâmetros foram explorados nos mesmos intervalos para todas as aplicações. Foram testadas populações de 50 a 100 indivíduos, taxas de cruzamento de 0,6 a 0,95 e taxas de mutação na faixa de 0,001 a 0,2. Existem alguns trabalhos recentes indicando formas de automação dos parâmetros do algoritmo, como Lobo (2000) e Lobo e Goldberg (2004), mas como este trabalho tem caráter exploratório, optou-se por examinar diversas combinações. Os melhores resultados ocorreram com populações de 100 indivíduos, taxas de cruzamento de 0,8 e de mutação de 0,008 (0,002 em um dos casos) e seleção elitista de 5% dos indivíduos. O algoritmo foi programado para parar em 100 iterações ou quando a razão entre o erro absoluto da melhor solução e o erro médio da população fosse menor do que 0,01 por mais de cinco gerações seguidas.

Os resultados indicaram equilíbrio entre os sistemas desenvolvidos. Para o conjunto de três regras, o erro estabilizou-se em 50 gerações, sendo atingido um nível de erro padrão com os dados de teste de 16.833 e de 18,11% nos erros absolutos percentuais. Para o sistema com cinco regras, os erros foram de EP=18.822 e EA=18,62%, respectivamente, sendo o conjunto de regras obtido em 95 gerações. Finalmente, o conjunto de sete regras foi definido em 65 gerações, e os erros foram de 17.718 e 18,35%. Em face do equilíbrio entre os conjuntos testados, a preferência deve recair no sistema mais simples, ou seja, o conjunto com o menor número de regras, por ter interpretação mais fácil, além de possuir efetivamente o menor erro entre os modelos estimados, no caso. O conjunto final, composto de três regras difusas, está apresentado na Figura 6. Os conjuntos difusos A_1, A_2 e A_3 são definidos com base nas áreas de 11,265 m^2, 236,326 m^2 e 454,652 m^2, respectivamente, sendo os dois extremos em formato trapezoidal e o central em formato triangular (ver Figura 5).

Os coeficientes e sinais dos modelos apresentados são coerentes com o esperado, não revelando surpresas. Os resultados estão apresentados na Tabela 3, indicando equilíbrio com os resultados da regressão, sendo este modelo difuso ligeiramente superior ao modelo de regressão.

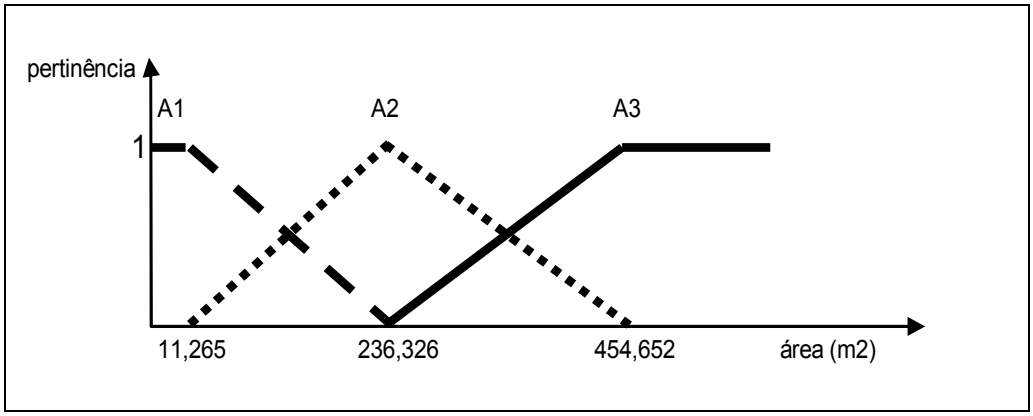

Figura 5 – **Conjuntos difusos para o sistema difuso baseado na área total com três regras**

Dados	EA	EP	Casos
Treino	17,30%	16.862	24.289
Teste	17,56%	16.980	6.074

Tabela 3 – **Resultados para as regras difusas extraídas da base de dados**

Sistema difuso com base na localização

A segunda alternativa de sistema difuso envolve a localização dos imóveis. Cada regra é "especializada" em uma determinada região da cidade, cujo centro de referência foi definido por análise de clusterização das coordenadas (X,Y) com os dados de treinamento. Foram inicialmente determinadas 10 regiões (correspondendo a 10 regras). O esquema adotado para o algoritmo genético foi semelhante ao do sistema baseado na área, exceto que as regras são individualmente apreciadas e a função de pertinência é espacial, contabilizando a distância de cada imóvel ao centro de referência da regra através das coordenadas dos imóveis (Equação 8):

$$Mi,j = 1 / (1 + [(Xi–Xj)2+(Yi–Yj)2]0,5.k) \quad (8)$$

Onde: Mi,j é o valor da pertinência do caso j à regra i, (X_i,Y_i) é o centro de referência da regra i, (X_j,Y_j) são as coordenadas de cada imóvel j, e k é um expoente, que permite variar a abrangência da função. O expoente k foi investigado, obtendo-se os melhores resultados com expoente unitário para todas as regras (k=1). As funções de pertinência empregadas definem na verdade regiões de pertinência no espaço urbano, obtidas pela giração (em 360^o) da curva exponencial $M_{i,j}$. Em função do formato dessa função, o valor de pertinência não é normalizado, ou seja, não está restrito ao intervalo [0,1], e depende da proximidade entre si dos centros das regras.

As regras foram examinadas utilizando-se a pertinência Mi,j como uma penalidade na função de fitness, para forçar o ajustamento localizado. Com essa finalidade, a função de ajuste tomou a seguinte forma (Equação 9):

$$Fi=1/(1+EAi'), \text{ com} \quad (9)$$
$$EAi'=\Sigma i,j\ [|Yj-Yhi,j| / Yj * 100 / Mi,j]$$

Onde: Fi é a medida do ajustamento da regra i, EAi' é a medida de erro percentual modificada, e os outros parâmetros são os já descritos acima. Com a aplicação desta penalidade, as regras mais ajustadas aos casos próximos ao centro da região são beneficiadas, pois os erros dos casos mais distantes (com pertinência pequena) são ampliados.

Na primeira fase de construção do sistema de regras (geração da base inicial), cada uma das 10 regras foi ajustada individualmente, com populações de 100 regras similares, inicialmente baseadas nas regras obtidas por regressão, aplicando variações aleatórias de ±50% em torno dos coeficientes da Equação 5. As regras foram ajustadas seguindo a ordem do número de casos contidos em cada *cluster* (ou seja, a regra R_1 corresponde ao grupo com maior quantidade de imóveis). Para facilitar a compreensão das regras, elas foram identificadas com as regiões da cidade correspondentes às regiões de pertinência maior (ver Tabela 4, regras R_1 a R_{10}).

Na segunda fase (refinamento da base), verificou-se o desempenho da base de regras inicial, identificando as regiões da cidade com os maiores erros e incluindo progressivamente novas regras para essas regiões. A base final contou com 15 regras, acrescentando-se novas regras para compensar erros localizados (no espaço) ou para determinados tamanhos de imóveis (regras R_{11} a R_{15}), apresentadas na Tabela 4. Com esse conjunto, os erros obtidos com o conjunto de dados de teste foram de 20.249 e 21,27%, para EP e erros percentuais (Tabela 5), ou seja, mesmo com o refinamento, os resultados finais são inferiores aos obtidos no modelo difuso anterior, baseado na área total. Para evitar repetições, não são apresentadas as 15 regras desse sistema de regras difusas. As equações têm formato similar às equações apresentadas na Figura 6.

R_1: SE Área é A_1 ENTÃO $Valor_1$=(363,399+38,225*Artocons+818,141*Razão +544,665* Med+1.016,166* Fin+2.107,615*Lux–24,088*Idade+14,932*Bairro–33,617*Centro– 94,569*Comércio+1.399,788*Vista–14,880*Mês)$^{4/3}$

R_2: SE Área é A_2 ENTÃO $Valor_2$=(1.147,260+33,432*Artocons+589,584*Razão +838,673* Med+ 2378,054* Fin+ 4.599,690*Lux–50,621*Idade+26,549*Bairro–118,912*Centro– 64,728*Comércio+1.917,465*Vista–28,918*Mês)$^{4/3}$

R_3: SE Área é A_3 ENTÃO $Valor_3$=(1.414,908+27,101*Artocons+5.593,242*Razão +4.565,306*Fin+7.516,682* Lux–43,623*Idade+ 49,287*Bairro–726,720*Centro–103,582*Mês)$^{4/3}$

Figura 6 – **Conjunto final de regras difusas**

Regra	Região	Centro	
		X	Y
R_1	Centro/Cidade Baixa	0,659	-0,673
R_2	Petrópolis	3,728	-1,570
R_3	São João/Higienópolis	3,597	1,708
R_4	Azenha	1,287	-3,136
R_5	Rubem Berta	11,566	2,063
R_6	Cristo Redentor	7,176	1,818
R_7	Cavalhada/Cristal	-0,678	-7,427
R_8	Bom Jesus	7,654	-1,903
R_9	Vila Nova	1,427	-10,996
R_{10}	Restinga	8,082	-13,574
R_{11}	Floresta/Moinhos de Vento (imóveis médios)	2,200	0,780
R_{12}	Bela Vista	4,000	-0,400
R_{13}	Humaitá	3,850	5,440
R_{14}	Moinhos de Vento/Mont'Serrat (imóveis grandes)	3,200	0,286
R_{15}	Centro	–0,528	-0,394

Tabela 4 – **Centros das funções de pertinência para o SBRDE baseado na localização**

O procedimento de acréscimo de regras poderia ser estendido, possivelmente com o aperfeiçoamento das estimativas, porém com incremento na complexidade e na dificuldade de compreensão da base. Outra forma de aperfeiçoamento é a subdivisão de cada uma das regras em conjuntos de regras especializados em faixas de áreas, adotando um sistema misto das duas alternativas apresentadas, com área e distância influindo simultaneamente no cálculo da pertinência às regras.

Dados	EA	EP	Casos
Treino	20,68%	19.853	24.289
Teste	21,27%	20.249	6.074

Tabela 5 – **Resultados do modelo difuso baseado na localização**

Comparação dos modelos

Foram calculados outros parâmetros para os três modelos. Além das medidas de erros indicadas acima, foram calculados o coeficiente de dispersão (COD), as correlações entre o preço original e o preço calculado pelas equações, e o número de erros menores do que 5% e 10% e maiores do que 50% (Tabela 6). Essas medidas foram calculadas empregando-se o arquivo com os dados de teste. Os modelos apresentaram bons resultados, em geral, embora com valores de COD relativamente altos nas três abordagens. A International Association of Assessing Officers (IAAO) recomenda um limite de 10% para modelos de avaliação em massa (IAAO, 1990). Esses resultados provavelmente estão associados com o tamanho e a diversidade da amostra. Por outro lado, os modelos têm altos coeficientes de correlação entre os preços reais e estimados ($r_{P,V}^{h}$~0,99), bem como um grande número de erros pequenos e pequeno número de erros grandes.

O pequeno número de erros grandes (potencialmente *outliers*) indica a importância do tratamento inicial (pré-processamento) dos dados. Também deve ser ressaltado o uso de diferentes arquivos de dados, originados da mesma amostra, mas divididos aleatoriamente para propósitos de treinamento e teste dos modelos. Esse procedimento aumenta a confiança nos modelos gerados, evitando *overfitting*[1].

Os modelos apresentados são similares, com uma pequena vantagem para os modelos difusos baseados no tamanho dos imóveis. Aparentemente, o desempenho de cada técnica depende parcialmente das características dos dados. Nesta aplicação, os dados empregados continham comportamentos aproximadamente lineares. Em outros casos, porém, as características dos dados poderão enfatizar o uso dos modelos difusos. Assim, é importante investigar técnicas alternativas com vistas a verificar os resultados em cada caso. Na abordagem difusa, uma das vantagens que podem ser observadas é a possibilidade do uso do conhecimento disponível como solução inicial para os algoritmos genéticos, tais como os modelos determinados por outras

[1] *Overfitting* é o excesso de ajustamento aos dados. Neste caso, o modelo inclui também o "ruído" dos dados (HAYKIN, 2001).

técnicas. As regras difusas também podem ser vistas como modelos hedônicos, mas compostas em um sistema de regras, ao contrário dos modelos de equação única tradicionais.

Aplicações

Os modelos calculados podem ser utilizados em diversas situações, tais como estudos de viabilidade e tributação. Uma das utilizações possíveis é ilustrada a seguir. Suponha-se que um empreendedor analise duas propostas de projetos para edifícios residenciais, um contando com 20 unidades de 100 m^2 cada (P1) e outro com 10 unidades de 200 m^2 cada (P2), em diferentes localizações na cidade. Se os custos de construção forem similares, a decisão dependerá dos preços de venda prováveis (estimados pelos modelos de mercado). Os atributos desses projetos são apresentados na Tabela 7.

Usando os modelos apresentados acima, os valores calculados para os projetos alternativos são os apresentados na Tabela 8. Para os três modelos, a alternativa P1 tem valores de mercado superiores aos da alternativa P2. Em vista desses resultados, o empreendedor pode concluir que o projeto P1 é a melhor alternativa, atingindo um valor de mercado cerca de 3% maior do que a outra alternativa, em média.

Modelo	COD	$r_{P.V}{}^{h}$	e<5%	e<10%	e>50%
Regressão	13,91	0,9898	1.244	2.392	95
Regras difusas baseadas na área	13,85	0,9898	1.286	2.416	87
Regras difusas baseadas na localização	14,04	0,9894	1.234	2.385	86

Tabela 6 – **Comparação geral entre os modelos**

Variável	P1	P2
Artocons	100	200
Razão	0,8	0,8
Pop	0	0
Med	0	0
Fin	1	1
Lux	0	0
Idade	0	0
Bairro	50	75
Centro	4,12	2,00
Comércio	0,50	1,00
Vista	0	0
Mês	37	37
Coordenada X	1,00	2,00
Coordenada Y	–4,00	0,00

Tabela 7 – **Características dos casos-exemplo (uma unidade)**

Modelo	P1–20x100	P2–10x200
Regressão	1.533.793	1.506.405
Regras difusas baseadas na área	1.554.780	1.496.335
Regras difusas baseadas na localização	1.543.355	1.494.025
Média	*1.543.975*	*1.498.923*

Tabela 8 – **Valores calculados para os casos-exemplo (total - para o prédio, em reais)**

Conclusão

A análise microeconômica do mercado local é uma fase importante na análise de viabilidade de novos empreendimentos. Os empreendedores precisam estimar o valor de mercado para os imóveis em estudo. Em muitos casos são utilizados processos subjetivos, aumentando-se os riscos do negócio. Estimativas mais precisas dos valores dos imóveis podem ser obtidas utilizando-se técnicas de avaliação em massa, através de modelos de regressão ou de técnicas de inteligência artificial, tais como as regras difusas. Os modelos são construídos com base em dados de mercado, e as estimativas das vendas futuras podem ser determinadas com maior precisão e assim alguns dos riscos envolvidos no desenvolvimento de novos empreendimentos podem ser reduzidos. Este trabalho discute a construção de modelos de avaliação em massa, utilizando regressão e regras difusas. Foram apresentadas noções de lógica difusa e de sistemas baseados em regras difusas. Foi apresentada uma aplicação com dados de vendas de apartamentos em Porto Alegre, comparando-se três modelos distintos. O primeiro segue a análise de regressão, no formato tradicional, enquanto os outros dois são sistemas de regras difusas, baseados no tamanho e na localização dos imóveis, ajustados com auxílio de algoritmos genéticos. Os três modelos obtidos são similares em formato e também nas medidas de erro. Os resultados demonstram que as duas técnicas podem gerar modelos adequados para análise do mercado, e a escolha da técnica depende de investigação com os dados reais, em cada caso.

Referências

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **Avaliação de bens -** Parte 2 - Avaliação de Imóveis Urbanos: NBR-14653-2. Rio de Janeiro: ABNT, 2004.

ALCALÁ, R.; CASILLAS, J.; CORDÓN, O.; HERRERA, F.; ZWIR, S. J. I. Techniques for learning and tuning fuzzy rule-based systems for linguistic modeling and their applications. In: LEONDES, C. T. (Ed.). **Knowledge-based systems**. Techniques and applications, v. 3. Academic Press, 2000. p. 889-941.

BAGNOLI, C.; SMITH, H. C. The theory of fuzzy logic and its application to real estate valuation. **Journal of Real Estate Research**, v. 16, n. 2, p. 169-199, 1998.

BONISSONE, P. P.; CHEETAM, W.; GOLIBERSUCH, D. C.; KHEDKAR, P. Automated residential property valuation: An accurate and reliable approach based on soft computing. In: RIBEIRO, R.; ZIMMERMANN, H.; YAGER, R. R.; KACPRZYK, J. (Ed.). **Soft computing in financial engineering**. Heidelberg: Physica-Verlag, 1998.

BONISSONE, P. P.; CHEN, Y-T.; GOEBEL, K.; KHEDKAR, P. S. Hybrid soft computing systems: Industrial and commercial applications. **Proceedings of the IEEE**, v. 87, n. 9, p. 1641-1667, Sept. 1999.

BYRNE, P. Fuzzy analysis: a vague way of dealing with uncertainty in real estate analysis? **Journal of Property Valuation and Investment**, v. 13, n. 3, p. 22-41, 1995.

CHAO, L.-C.; SKIBNIEWISKI, M. J. Fuzzy logic for evaluating alternative construction technology. **Journal of Construction Engineering and Management**, v. 124, n. 4, p. 297-304, July 1998.

CIOS, K. J.; PEDRYCZ, W.; SWINIARSKI, R. W. **Data mining**: methods for knowledge discovery. Boston: Kluwer, 1998.

CORDÓN, O.; HERRERA, F. A two-stage evolucionary process of designing TSK fuzzy rule-based systems. **IEEE Transactions on Systems, Man, and Cybernetics** - Part B: Cybernetics, v. 29, n. 6, 1999.

CORDÓN, O.; HERRERA, F.; HOFFMANN, F.; MAGDALENA, L. **Genetic fuzzy systems**: Evolutionary tuning and learning of fuzzy knowledge bases. World Scientific: Singapore, 2001.

GOLDBERG, D. E. **Genetic algorithms in search, optimization, and machine learning**. Reading: Addison-Wesley, 1989.

GÓMEZ-SKARMETA, A. F.; JIMÉNEZ, F. Generating and tuning fuzzy rules using hybrid systems. In: FUZZ/IEEE'97, 1997, Barcelona. **Proceedings...** Piscataway (NJ): IEEE, 1997. p. 247-252.

GONZÁLEZ, M. A. S. **Aplicação de descobrimento de conhecimento em bases de dados e inteligência artificial em avaliação de imóveis**. 2002. Tese (Doutorado em Engenharia Civil) - Universidade Federal do Rio Grande do Sul, Porto Alegre.

GONZÁLEZ, M. A. S. **Metodologia de avaliação de imóveis**. Novo Hamburgo: SGE, 2003.

GUJARATI, D. N. **Econometria básica**. São Paulo: Makron, 2000.

HARVEY, J. **Urban land economics**. 4. ed. London: MacMillan, 1996.

HAYKIN, S. **Redes neurais**: fundamentos. Porto Alegre: Artmed, 2001.

IAAO (INTERNATIONAL ASSOCIATION OF ASSESSING OFFICERS). **Standards on ratio studies**. Chicago: IAAO, 1990.

KACPRZYK, J. **Multistage fuzzy control**. A model-based approach to fuzzy control and decision making. Chichester: John Wiley, 1997.

KARRAY, F.; ZANELDIN, E.; HEGAZY, Y.; SHABEEB, A. H. M.; ELBELTAGY, E. Tools of soft computing as applied to the problem of facilities layout planning. **IEEE Transactions on Fuzzy Systems**, v. 8, n. 4, p. 367-379, Aug. 2000.

KOHAVI, R.; JOHN, G. H. Wrappers for feature subset selection. **Artificial Intelligence**, v. 97, n. 1-2, p. 273-324, 1997.

LAVENDER, S. D. **Economics for builders and surveyors**. Essex (UK): Longman, 1990.

LEU, S-S.; CHEN, A-T.; YANG, C-H. Fuzzy optimal model for resource-constrained construction scheduling. **Journal of Computing in Civil Engineering**, v. 13, n. 3, p. 207-216, July 1999.

LOBO, F.G. **The parameter-less genetic algorithm**: rational and automated parameter selection for simplified genetic algorithm operation. 2000. Tese (Doutorado) - Universidade Nova de Lisboa, Lisboa.

LOBO, F. G.; GOLDBERG, D. E. The parameter-less genetic algorithm in practice. **Information Sciences**, v. 167, n. 1-4, p. 217-232, Dec. 2004.

NGUYEN, H. T.; WALKER, E. A. **A first course in fuzzy logic**. 2. ed. Boca Ratón: Chapman and Hall, 2000.

ROBINSON, R. **Housing economics and public policy**. London: MacMillan, 1979.

SHEPPARD, S. Hedonic analysis of housing markets. In: CHESHIRE, P. C.; MILLS, E. S. (Ed.). **Handbook of applied urban economics**, v. 3. New York: Elsevier, 1999, chap. 8.

SOH, C. K.; YANG, J. Fuzzy controlled genetic algorithm search for shape optimization. **Journal of Computing in Civil Engineering**, v. 10, n. 2, p. 143-214, Apr. 1996.

SOIBELMAN, L.; GONZÁLEZ, M. A. S. A Knowledge Discovery in Databases Framework to Property Valuation. **Journal of Property Tax Assessment and Administration**, v. 7, n. 2, p. 77-106, 2002.

YANG, Y.; SOH, C. K. Fuzzy logic integrated genetic programming for optimization and design. **Journal of Computing in Civil Engineering**, v. 14, n. 4, p. 249-254, Oct. 2000.

ZADEH, L. A. Fuzzy sets. **Information and control**, v. 8, p. 338-352, 1965.