



UNIVERSIDADE ESTADUAL DO CEARÁ

CARLOS CRISTIANO CABRAL

**EMPREGO DE TÉCNICAS DE CLUSTERIZAÇÃO
PARA GESTÃO TRIBUTÁRIA DO SOLO URBANO:
UM ESTUDO DE CASO APLICADO À CIDADE DE
FORTALEZA**

**FORTALEZA – CEARÁ
2010**

CARLOS CRISTIANO CABRAL

EMPREGO DE TÉCNICAS DE CLUSTERIZAÇÃO PARA GESTÃO TRIBUTÁRIA
DO SOLO URBANO: UM ESTUDO DE CASO APLICADO À CIDADE DE
FORTALEZA

Dissertação apresentada à Coordenação do Mestrado Acadêmico em Ciência da Computação do Centro de Ciência e Tecnologia da Universidade Estadual do Ceará, como requisito para a obtenção do título de Mestre em Ciência da Computação.

Área de Concentração: Sistemas de Apoio à Decisão.

Orientador: Prof. Dr. Jackson Sávio de Vasconcelos Silva.

FORTALEZA – CEARÁ
2010

C117e Cabral, Carlos Cristiano

Emprego de técnicas de clusterização para gestão tributária do solo urbano: um estudo de caso aplicado à cidade de Fortaleza / Carlos Cristiano Cabral . — Fortaleza, 2010.

xvii, 136 p. : il. ; 18cm.

Orientador: Prof. Dr. Jackson Sávio de Vasconcelos Silva.

Dissertação (Mestrado Acadêmico em Ciência da Computação) – Universidade Estadual do Ceará, UECE, Centro de Ciência e Tecnologia, MACC, 2010.

Referências bibliográficas: p. 128-131.

1. Clusterização. 2. Redes Neurais. 3. Aprendizado. 4. Mapas auto-organizáveis de Kohonen. 5. IPTU. 6. Cadastro Imobiliário. I. Universidade Estadual do Ceará, Centro de Ciência e Tecnologia, MACC.

CDD: 001.6

CARLOS CRISTIANO CABRAL

EMPREGO DE TÉCNICAS DE CLUSTERIZAÇÃO PARA GESTÃO TRIBUTÁRIA
DO SOLO URBANO: UM ESTUDO DE CASO APLICADO À CIDADE DE
FORTALEZA

Dissertação apresentada à Coordenação do Mestrado Acadêmico em Ciência da Computação do Centro de Ciência e Tecnologia da Universidade Estadual do Ceará, como requisito para a obtenção do título de Mestre em Ciência da Computação.

Área de Concentração: Sistemas de Apoio à Decisão.

Data da aprovação: 31/05/2010

BANCA EXAMINADORA

Prof. Dr. Jackson Sávio de Vasconcelos Silva (Orientador)
Universidade Estadual do Ceará – UECE

Prof. Dr. Antônio Clécio Fontelles Thomaz
Universidade Estadual do Ceará – UECE

Profa. Dra. Gilda Collet Bruna
Universidade Presbiteriana Makenzie – UPM

Prof. MSc. Alexandre Sobreira Cialdini
Universidade Federal do Ceará – UFC

Dedico esta dissertação a meus amados pais Vera (*in memoriam*) e Cabral, aos meus irmãos Sidney e Marcelo, à minha esposa Márcia e aos meus filhos Daniel e Maria Clara, exemplos perenes de unidade familiar e de amor fraterno.

AGRADECIMENTOS

Devoto meus mais profundos e sinceros agradecimentos:

A Deus, em primeiro lugar, pela bondade infinita em me deixar provar do Seu amor e por me dar a graça de amar e ser amado, de ser feliz com o que faço e agradecido pelo que tenho.

Ao meu orientador e antigo colega de magistério, Prof. Dr. Jackson Sávio de Vasconcelos Silva, que, além de me convencer de que seria possível conciliar minhas atividades laborais diárias com a dedicação ao mestrado, usou de inimaginável paciência e compreensão diante de minhas limitações pessoais e profissionais, viabilizando a conclusão deste trabalho.

À Secretaria de Finanças do Município de Fortaleza, em especial ao Ilmo. Sr. Secretário Alexandre Sobreira Cialdini, cujo apoio *incontinenti* e contínuo foram fundamentais, para que conseguisse percorrer, ileso, o longo caminho até a obtenção deste valoroso título acadêmico. Também à minha Coordenadora de Tributos, Sra. Ivany Araújo, pelo apoio e tolerância com que me brindou nos momentos de ocasional ausência física ou psicológica em detrimento de minhas responsabilidades acadêmicas.

Aos colegas do Mestrado Acadêmico em Ciência da Computação (MACC), na pessoa do colega Enyo Gonçalves, pelo exemplo de dedicação e amor às atividades acadêmicas.

Ao amigo Marco “Cinzento”, pela disponibilidade e persistência com que me auxiliou, inicial e expressivamente, com a clusterização, fornecendo-me material de estudo e desmistificando o funcionamento das Redes Neurais, sugerindo ainda muitos melhoramentos no *software* por mim desenvolvido, para exemplificar a aplicação prática da teoria apresentada nesta dissertação.

Aos funcionários da secretaria do MACC, especialmente aos senhores Marcos Ferreira Lima e Wagner Souza, que sempre agiram com distinta cortesia e atenção, propiciando um ambiente tranquilo e favorável para minhas pesquisas e estudos, tanto quanto à Biblioteca Central da UECE, na pessoa da Sra. Thelma Melo e do Sr. Welton Rios, pela total disponibilidade e cortesia na revisão da diagramação e da bibliografia desta dissertação.

A todos os professores do MACC, especialmente aos doutores Gustavo Augusto Lima de Campos, Antônio Clécio Fontelles Thomaz e Renato Craveiro pelo contínuo, contagiante e sempre inovador estímulo às minhas conquistas e desafios, convencendo-me da utilidade prática deste trabalho.

À professora Gilda Collet Bruna, pelos seus valiosos comentários e sugestões para o incremento e esclarecimento do texto desta dissertação, conferindo a ele maior clareza e abrangência, facilitando seu uso e compreensão por profissionais de áreas diversas da Computação e do Direito.

Ao colega de trabalho, Auditor do Tesouro Municipal, Valdir Machado Calixto, pela presteza na leitura dos capítulos iniciais da dissertação, sugerindo correções

ortográficas e gramaticais, e aconselhando sobre assuntos adicionais a serem abordados, com visão tributária.

Aos meus irmãos Sidney e Marcelo, bem como ao tio Jesus, pela pronta disponibilidade em ler, analisar e ponderar sobre este texto, conferindo-lhe visão crítica e também revendo muitos termos e enfoques além da ortografia e da própria estrutura. Isto é ser família!

À Márcia, esposa, amiga, companheira e cúmplice, cuja fé, compreensão, apoio e dedicação à família foram extremamente importantes nos momentos em que não me podia faltar concentração para os estudos, só Deus sabe o quão longe ainda chegaremos por formarmos, juntos, um time de verdade, uma equipe campeã.

Ao Daniel e à Maria Clara, por toda a sensibilidade e entendimento de que o papai precisava privá-los de tantos momentos de carinho, companhia e lazer, em prol de um bem maior. Sei o quão sinceras foram todas as suas orações pedindo a Deus e a Nossa Senhora pelo sucesso deste mestrado.

Ao meu pai, Cabral, devo-lhe gratidão e reconhecimento eternos. Pelo exemplo de dedicação e entrega à família, pela firmeza e serenidade para me guiar nas mais difíceis decisões e, especialmente, por me incentivar a novos desafios que entendia quase intransponíveis. Amo muito você, meu pai, e sei que a mamãe, mesmo já tendo ido ao encontro do Senhor, jamais deixou de interceder por nós.

RESUMO

O presente trabalho busca empregar técnicas de clusterização para identificar distorções tributárias causadoras de evasão fiscal, fornecendo embasamento para a correção, o controle e a minimização da ocorrência de inconsistências em bancos de dados técnicos imobiliários. Estudadas algumas redes neurais, a Rede Neural de Kohonen, também conhecida como Mapas Auto-organizáveis de Kohonen, foi escolhida por tratar com maior habilidade os dados relativos aos imóveis, uma vez que estes tipos de dados apresentam grande duplicidade de valores em suas características. Apresenta-se, a título de avaliação, estudo de caso específico do município de Fortaleza. A rede neural implementada deu origem ao software Analisador do Solo Urbano, atualmente na versão 1.4.1.

Palavras-chave: clusterização, redes neurais, Kohonen, mapas auto-organizáveis, tributação imobiliária, IPTU, arrecadação, propriedade urbana.

ABSTRACT

The present work aims to employ clustering technics to identify tax distortions that cause fiscal loss, serving as the basis for the correction, control and minimization of the inconsistencies in the real estate technical databases. Checked some neural networks, the Kohonen Neural Network also known as the Kohonen Self-Organizing Maps was chosen because treats real estate data with better ability than others, since these sort of data present large value duplication in their characteristics. It is presented, as an evaluation, a specific case study on the city of Fortaleza. The neural network implemented led to the software Analisador do Solo Urbano, currently in version 1.4.

Keywords: clustering, neural networks, Kohonen, self-organizing maps, real estate tax, IPTU, collection, urban property.

Sumário

RESUMO.....	VIII
ABSTRACT.....	IX
LISTA DE FIGURAS	XIII
LISTA DE QUADROS.....	XIV
LISTA DE TABELAS	XV
LISTA DE EQUAÇÕES	XVI
LISTA DE ABREVIATURAS.....	XVII
1 INTRODUÇÃO	18
2 CONCEITOS TRIBUTÁRIOS.....	32
2.1 Considerações iniciais	32
2.2 Elementos da tributação imobiliária	33
2.3 O cadastro imobiliário	35
2.3.1 Características do cadastro	35
2.3.2 Atualização cadastral.....	36
2.3.3 Riscos nas avaliações em massa.....	39
2.3.4 Distribuição espacial	39
2.4 Considerações finais.....	42
3 FUNDAMENTAÇÃO TEÓRICA	44
3.1 Considerações iniciais	44
3.2 A estrutura cerebral.....	44
3.3 Redes Neurais Artificiais (RNA)	47
3.4 Aprendizado	49
3.4.1 Aprendizado supervisionado.....	51
3.4.2 Aprendizado não-supervisionado	52
3.5 Normalização de dados.....	54

3.6	Características dos dados	58
3.7	Medida de similaridade	60
3.8	Vizinhança	62
3.9	Critérios de parada	64
3.10	Os mapas auto-organizáveis de Kohonen	65
3.11	Considerações finais	73
4	ESTUDO DE CASO	75
4.1	Considerações iniciais	75
4.2	Equipamento utilizado	76
4.3	A ferramenta de apoio à decisão	77
4.4	Criação dos ambientes gráficos	80
4.4.1	Ambiente gráfico de mapeamento da pesquisa	81
4.4.2	Ambiente gráfico de competição	84
4.4.3	Ambiente gráfico de classificação	85
4.5	Obtenção de dados brutos	86
4.5.1	Conversão do padrão da base de dados	87
4.5.2	Elaboração das bases de dados auxiliares	88
4.5.3	Identificação e dimensionamento dos campos	89
4.5.4	Análise de dados	90
4.5.4.1	<i>Relatório GERAL por distrito</i>	91
4.5.4.2	<i>Relatório GERAL por atributo</i>	92
4.5.4.3	<i>Relatório GERAL por atributo (planilha)</i>	93
4.5.4.4	<i>Relatório de inconsistências</i>	94
4.5.4.5	<i>Relatório de anormalidades</i>	95
4.6	Definição da pesquisa	95
4.6.1	Seleção da área exploratória	96
4.6.2	Seleção dos parâmetros de pesquisa	97
4.6.3	Pesquisa da composição imobiliária	99
4.7	Experimento de treinamento e classificação	102
4.7.1	Configuração da Camada de Entrada	103
4.7.2	Configuração da Camada de Saída	104
4.7.3	Definição dos critérios de parada	105
4.7.4	Normalização dos dados	109
4.7.5	Treinamento da rede	112
4.7.6	Classificação	113
4.8	Sugestão de atividades corretivas para aumento de arrecadação	116
5	TRABALHOS CORRELATOS	119

5.1	Considerações iniciais	119
5.2	Revisão bibliográfica	119
5.2.1	Trabalho 1: “Análise da utilização de inteligência artificial na determinação de plantas de valores”	119
5.2.2	Trabalho 2: “Da teoria à prática - introduzindo técnicas emergentes na avaliação de imóveis em massa”	121
5.2.3	Outros trabalhos	123
5.3	Considerações finais	124
6	CONCLUSÕES	125
	BIBLIOGRAFIA	128
	GLOSSÁRIO	132
	APÊNDICE A – LISTAGEM DA PESQUISA DE COMPOSIÇÃO IMOBILIÁRIA ...	142
	APÊNDICE B – TREINAMENTO (AJUSTE DE PESOS) DA RNA	145
	ANEXO A – TABELA DE CONVERSÃO DE DISTRITO PARA BAIRRO	149
	ANEXO B – CÁLCULO DO IPTU	150
	ANEXO C – TABELA DE VALORES DE EDIFICAÇÃO	153

LISTA DE FIGURAS

Figura 1 – Divisão do município de Fortaleza em Secretarias Executivas Regionais (SER).....	21
Figura 2 – Divisão do município de Fortaleza em distritos	40
Figura 3 – Subdivisão do distrito 1 (parte do bairro Centro) e suas quadras	41
Figura 4 – Estrutura elementar de um neurônio cerebral	45
Figura 5 – Rede com associações lineares em duas camadas.....	47
Figura 6 – Estrutura de relacionamento entre neurônios.....	48
Figura 7 – Exemplo de normalização de dados para variáveis contínuas e discretas	58
Figura 8 – Representação espacial das características do banco de dados	60
Figura 9 – Visualização da Distância Euclidiana (d) em plano bi-dimensional.....	61
Figura 10 – Exemplos de topologias de vizinhança do neurônio vencedor.....	62
Figura 11 – Exemplos de topologias das Camadas de Saída	66
Figura 12 – Fluxograma do funcionamento da Rede Neural de Kohonen	69
Figura 13 – Especialização da RNA com base na quantidade de neurônios da Camada de Saída....	70
Figura 14 – Tela inicial do <i>software</i> Analisador do Solo Urbano em sua versão 1.4.1.	78
Figura 15 – Concentração de imóveis em geral no distrito 15 (Aldeota)	81
Figura 16 – Concentração de imóveis comerciais no distrito 15 (Aldeota).....	82
Figura 17 – Exemplos de mapeamento do neurônio vencedor e de seus vizinhos	85
Figura 18 – Agrupamentos formados após a classificação das entradas	86
Figura 19 – Imóveis com apartamentos de cobertura nos distritos	94
Figura 20 – Pesquisa de composição imobiliária através da análise do distrito.....	101
Figura 21 – Ocorrências de imóveis com áreas de terreno maiores que 1.000 m ² no distrito 15	101
Figura 22 – Legenda sobre a quantidade de imóveis identificados na quadra	102
Figura 23 – Seleção de neurônios para a composição da Camada de Entrada	103
Figura 24 – Camada de saída, também chamada de Camada de Competição.....	104
Figura 25 – Critérios de parada adotados pelo Analisador.....	106
Figura 26 – Desaceleração na aprendizagem com o tempo	107
Figura 27 – Configuração dos parâmetros de treinamento no Analisador	109
Figura 28 – Definição de classes para a normalização das entradas	109
Figura 29 – Ambiente de treinamento da Rede Neural de Kohonen.....	114
Figura 30 – Relação entre os neurônios de saída e os agrupamentos formados	115
Figura 31 – Agrupamentos formados após a classificação do distrito 15	117

LISTA DE QUADROS

Quadro 1 – Variação na alíquota do IPTU do município de Fortaleza.	34
Quadro 2 – Exemplo de cartografia tradicional.....	42
Quadro 3 – Especificação do equipamento utilizado no estudo de casos.....	77
Quadro 4 – Fontes de dados utilizados no estudo de casos	87
Quadro 5 – Inconsistências ou suspeições detectadas durante a análise do Cadastro	95
Quadro 6 – Anormalidades causadoras de exclusão de imóveis do cadastro Imobiliário.....	96
Quadro 7 – Normalização de variável contínua	111
Quadro 8 – Normalização de variável discreta	112
Quadro 9 – Parametrização da Rede Neural para treinamento.	112
Quadro 10 – Quadro comparativo de performance para treinamento e classificação.	114
Quadro 11 – Quantitativo de neurônios nos agrupamentos, após a classificação do distrito 15	116

LISTA DE TABELAS

Tabela 1 – Subdivisão em classes dos dados contínuos de uma amostra ($lc \rightarrow Lc$)	57
Tabela 2 – Subdivisão em classes dos dados discretos de uma amostra	58
Tabela 3 – Exemplo de dimensões (características) em um banco de dados imobiliário	59
Tabela 4 – Grau de especialização em função da dimensão da Camada de Saída.....	71
Tabela 5 – Comparação quantitativa do distrito 15 com o município	100

LISTA DE EQUAÇÕES

Equação 1 – Erro $e(t)$ entre a saída calculada e a saída desejada	51
Equação 2 – Ajuste de pesos entre as conexões	52
Equação 3 – Ajuste de peso do neurônio vencedor e de seus vizinhos	53
Equação 4 – Interpolação linear pelo método Máx-Min.....	55
Equação 5 – Amplitude total de uma amostra	56
Equação 6 – Amplitude do intervalo de classes	56
Equação 7 – Limite inferior da classe c	56
Equação 8 – Limite superior da classe c	57
Equação 9 – Distância euclidiana entre dois vetores X e Y	61

LISTA DE ABREVIATURAS

ARM	Análise de Regressão Múltipla
CELTI	Célula de Tecnologia da Informação
CGIPTU	Célula de Gestão do IPTU
CSV	<i>Comma Separated Values</i>
DAS	Direção, Chefia e Assessoramento
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IPTU	Imposto sobre a Propriedade Predial e Territorial Urbana
MLP	<i>Multilayer Perceptron</i>
PGVI	Planta Genérica de Valores Imobiliários
RBC	Raciocínios baseados em casos
RNA	Rede Neural Artificial
SEFIN	Secretaria de Finanças do Município de Fortaleza
SER	Secretaria Executiva Regional
SIG	Sistemas de Informações Geográficas
SOM	<i>Self-Organizing Maps</i>
SQL	<i>Structured Query Language</i>
TXT	Formato texto padrão

1 INTRODUÇÃO

Vive-se em um mundo onde a tecnologia se apresenta a todo instante a serviço do cidadão, seja através de equipamentos ou através de metodologias que oferecem a possibilidade de agregar mais eficiência àquilo que se faz. Tal característica é especialmente observada na Administração Pública, quando o Administrador Público precisa ser eficiente, produzindo o efeito desejado e dando bons resultados com o seu trabalho, como assevera Moraes (2002, p.303). A eficiência é princípio mandatário estatuído no *caput* do artigo 37 da Constituição Federal de 1988, conforme Pinto, Windt e Céspedes (2008, p.25).

Como destaca Alexandre (2008, p.203), o exercício da competência tributária, em sentido amplo, engloba a instituição, a arrecadação, a fiscalização e a execução de leis, serviços, atos ou decisões administrativas, sendo do município a competência tributária sobre diversos impostos, dentre eles o Imposto sobre a Propriedade Predial e Territorial Urbana (IPTU), cuja arrecadação servirá para financiar a máquina pública e para oferecer serviços à população, tais como educação, saúde, transporte, cultura, lazer e tantos outros. Para que o IPTU possa ser cobrado, é necessário que a Administração Tributária do município seja capaz de aferir o valor de cada imóvel, pois é a partir dele que é calculado o valor do imposto a ser pago. Ora, a incorreta aferição do valor do imóvel, na predominância das vezes, leva a vultosos danos para a arrecadação municipal e, conseqüentemente, para os serviços a serem revertidos para a população. Ocasionalmente leva, ainda, a uma avaliação superestimada do imposto devido, com tributação excessiva para o contribuinte. Mas como poderiam ser minimizados estes danos?

Um imóvel tem seu valor baseado em sua localização, tanto quanto em suas características como, tamanho do terreno, área construída, idade, estado de conservação, material utilizado na construção e tantas outras que o próprio mercado imobiliário dita como agregadoras ou redutoras de valor. O cadastro de imóveis mantido pelos órgãos tributantes com o objetivo de proporcionar o exercício da competência tributária deve, portanto, ser mantido atualizado, para que o cálculo do valor do imóvel e o respectivo imposto a pagar sejam corretamente computados. Seria tarefa simples, se não fossem a imensa quantidade de imóveis cadastrados e a enorme gama de características necessárias, para se calcular o valor de cada um deles.

Como a valorização dos imóveis é extremamente dinâmica em virtude da alteração de uma ou mais das características já mencionadas e existe comumente escassez de recursos tecnológicos, materiais, humanos e metodológicos, a atualização perene e tempestiva do Cadastro Imobiliário resta deveras dificultada. A desatualização traz inúmeras incorreções na tributação, ineficiência na fiscalização e prejuízos na arrecadação, especialmente porque o contribuinte do imposto, aquele que deve recolhê-lo aos cofres públicos, não atende espontaneamente à obrigação que normalmente a lei municipal lhe impõe, qual seja, a de manter o cadastro atualizado sempre que acontecer alteração nas características de seu imóvel, tal como reza o inciso IV do artigo 15 da Consolidação da Legislação Tributária do Município de Fortaleza, conforme Gomes e Mesquita (2003, p.24) .

Desta forma, à vista da desatualização cadastral, os equívocos tributários tornam-se um grave problema a ser solucionado ou, pelo menos, minimizado. Como então identificar, em todo o universo do Cadastro Imobiliário, aqueles imóveis cujo cadastro está provavelmente desatualizado? Seria razoável e, mais que isso, eficiente, pesquisar os imóveis um de cada vez através de verificação em campo e de detalhada análise cadastral, ou seria preferível usar um procedimento que identificasse muitos imóveis de uma só vez? Qual caminho seguir? Se a Administração Tributária tiver uma força de trabalho da ordem de 10^3 imóveis para cada servidor envolvido com a tributação do IPTU, já é razoável pensar em um mecanismo de identificação em massa. Se esta relação for ainda mais expressiva, como é o caso de várias grandes cidades brasileiras, onde a relação atinge valores mais elevados, é de bom alvitre pensar que a identificação em massa e automatizada é largamente aconselhada.

A análise de dados relativos ao cadastro, com cruzamento de informações e execução de procedimentos de detecção de inconsistências, listando e corrigindo equívocos cadastrais, é providência elementar e de fácil realização através de um programa de computador ou mesmo por meio de uma planilha eletrônica bem elaborada. São tarefas meramente lógicas, inferidas por dedução ou indução, quando, no primeiro caso, a conclusão é consequência necessária das premissas e, no segundo caso, a conclusão é consequência plausível das mesmas premissas, tal como ensinam Imaguire e Barroso (2006, p.30). São, portanto, atingíveis com um bom planejamento e com um algoritmo bem elaborado por alguém que conheça o panorama das regras de negócio, ou seja, com uma boa e clara sequência de

pressupostos intermediários a serem implementados, definindo os aspectos funcionais representativos do trabalho a ser realizado.

No entanto é preciso ir além. É preciso usar a tecnologia em relação aos procedimentos aplicando algum nível de raciocínio não absolutamente lógico. Segundo Beale e Jackson (1990, p.12) é algo como acontece com o cérebro, que é capaz de aprender com seus próprios erros e acertos. Com nível de falha aceitável, é preciso criar um mecanismo através do qual se possa perceber prováveis vícios cadastrais em um imóvel analisado para, enfim, serem atualizados os seus dados. A atualização deverá ser procedida com a devida atenção para cada caso, realizando-a em massa tanto quanto possível, em virtude da grande quantidade de dados tratados. Há que se alertar, entretanto, que atualizações em grande escala, quando inapropriadas, geram grandes transtornos com indesejáveis efeitos negativos tanto sociais quanto políticos, além do próprio custo administrativo para a reversão do problema causado. De qualquer forma, localizar as incorreções em um espectro de centenas de milhares de imóveis, já terá sido um grande avanço, ainda que para retificações manuais e paulatinas.

Há seis décadas foram iniciados os estudos que hoje culminaram com a aplicação das chamadas **Redes Neurais Artificiais**, que nada mais são do que estruturas baseadas no funcionamento dos neurônios cerebrais e que emprestam uma linha de raciocínio não lógico ao tratamento de informações com os mais diversos fins. A lógica, segundo Imaguire e Barroso (2006, p.12) não nos dá o conhecimento da verdade, mas apenas indica os caminhos que levam de uma verdade a outra, partindo de premissas e chegando a conclusões. Ao tratar a solução de problemas de forma diversa da lógica do tradicional “se → então”, as redes revelam inúmeras possibilidades de se desvendar situações notadamente complexas. Resolveu-se, assim, aplicar e implementar uma rede neural com base no modelo desenvolvido pelo finlandês Teuvo Kohonen, para a identificação das distorções de um cadastro imobiliário. Vale lembrar que a metodologia descrita neste trabalho é perfeitamente aplicável para outros cadastros, tributários ou não, com as devidas adaptações.

1.1 Problemas identificados

Para fins administrativos, o município de Fortaleza encontra-se subdividido em 6 regiões administrativas básicas e uma extraordinária, conhecidas como *Secretarias Executivas Regionais* (SER), tal como ilustrado na Figura 1, valendo esclarecer que a Secretaria Extraordinária do Centro fica localizada dentro do perímetro da Secretaria Executiva Regional II. No total, tais Regionais contêm, juntas, em sua área territorial, 77 distritos. Analisando dados fornecidos pelo Fisco municipal em 07/05/2010, apura-se que um distrito chega a possuir até 1.491 quadras e até 25.119 inscrições de imóveis. Tais distritos contêm, ainda, um total aproximado de 85 mil faces de quadras, com valor de tributação diferenciado para cada face. Os lotes nos quais uma quadra está dividida podem possuir um único imóvel, mas por vezes podem possuir dezenas deles, como no caso de condomínios verticais, residenciais ou comerciais, por exemplo. No total, o município de Fortaleza possui 656.571 unidades imobiliárias inscritas para fins tributários.

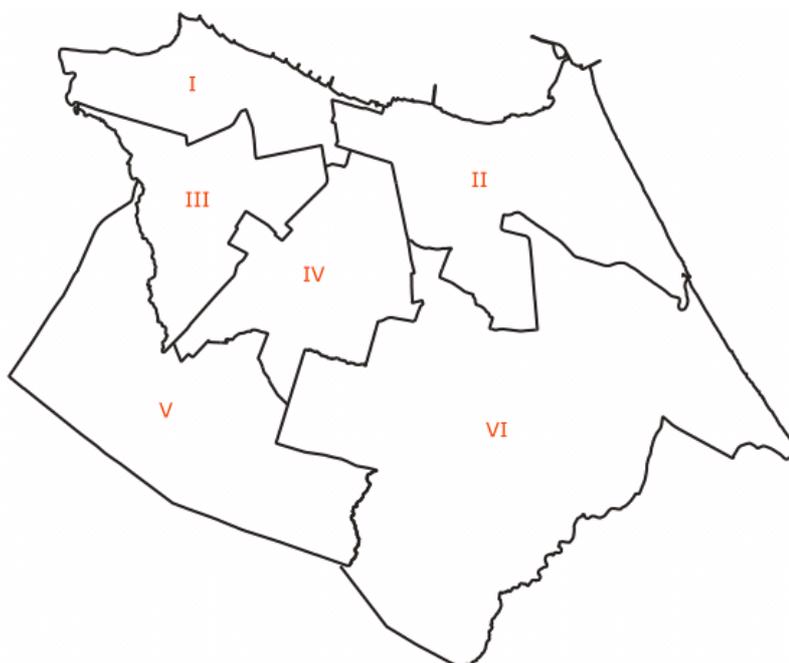


Figura 1 – Divisão do município de Fortaleza em Secretarias Executivas Regionais (SER)

A Secretaria de Finanças do Município, a quem cabe a administração tributária municipal, possui, em 07/05/2010, um quadro de 19 Auditores do Tesouro Municipal e 51 Assistentes Técnicos do Tesouro Municipal efetivamente envolvidos com o IPTU, além de estagiários e outros, quer lotados nos setores de análise de

processos, de atendimento, de auditoria, quer em cargos de direção, chefia e assessoramento (DAS), responsáveis direta ou indiretamente pela atualização cadastral dos imóveis, conforme dados da própria Secretaria.

Comumente, a identificação de distorções cadastrais é observada e trabalhada pontualmente ao longo das mais diversas atividades desenvolvidas pelos setores anteriormente elencados, onde também são implementadas ações localizadas e específicas de retificação cadastral em massa, a partir de inconsistências identificadas em procedimentos de auditoria interna. Todavia, a identificação e a correção findam, sobremaneira, restringidas e desaceleradas em virtude do baixo valor tecnológico agregado às metodologias empregadas.

Mesmo com fundamento em informações e relatórios gerenciais extraídos das bases de dados tributárias, o planejamento e a execução de ações tendentes à retificação cadastral são lentos, onerosos e de difícil implementação. Mas a Administração Pública, que deve observar suas prerrogativas e também os direitos dos administrados, como ensina Mello (2009, p.56), tem-se modernizado e não só tecnologicamente, mas, inclusive, quanto à qualificação de seus profissionais, o que tem viabilizado a adequação de novos e úteis mecanismos que permitem melhorar a prestação de serviços aos cidadãos.

Observa-se, adicionalmente, um vasto campo para a aplicação de metodologia científica, decorrente da demanda por um mecanismo que identifique com razoável facilidade imóveis geradores de evasão fiscal, quando da apuração de seus valores venais para fins da cobrança do IPTU, a fim de que sejam eles priorizados nas ações de fiscalização e retificação cadastral, tanto quanto em decisões para fins de tributação e arrecadação. Tal demanda advém, em geral, da grande complexidade da tributação e do grande volume de dados imobiliários tratados.

1.2 Hipótese

Quando tecnologia e metodologia científica especializadas são agregadas à identificação de distorções em bancos de dados tributários, independente de que espécies forem, tem-se a imediata possibilidade de melhoria da arrecadação para a Administração Tributária, especialmente diante da discricionariedade, citada por Mello (2009, p.426), que faculta ao Administrador Público a criação de

procedimentos específicos de trabalho, dentro dos ditames legais, com vistas a um melhor desempenho de suas atribuições,

A hipótese a ser testada, no caso específico do IPTU, é que a criação e a utilização de um mecanismo semi-automático de identificação de distorções imobiliárias se apresenta como solução para recuperação fiscal, permitindo ações precisas de retificação cadastral e fiscalização, bem como a adoção das corretas providências na legislação, dentre outras, maximizando-se a eficiência da Administração, diminuindo seus custos e minimizando as perdas de arrecadação.

A metodologia proposta envolve a apresentação dos imóveis pertencentes a uma determinada área geográfica como elementos de entrada em um modelo de rede neural denominada Mapas Auto-organizáveis de Kohonen. Através de um processo chamado de **treinamento**, ou aprendizado, a rede aprende a partir de seu ambiente, e melhora seu desempenho, como ensina Haykin (2001, p.75). Com o acúmulo de conhecimento sobre os dados apresentados, a rede classifica os imóveis que entende semelhantes, criando assim *clusters*, ou seja, agrupamentos padronizados nos quais os *clusters* se “corporificam”, assumindo identidade própria.

A partir da parametrização da rede, feita por um profissional a quem se convencionou chamar de **Operador da Rede**, e que domine as regras de negócio, qual seja, o cadastro imobiliário, o treinamento especializa-se ao ponto de dotar a rede de suficiente conhecimento para classificar não apenas os imóveis utilizados no próprio treinamento, mas qualquer outro que se lhe apresente. Eis, portanto, a abrangência deste trabalho, cujo objetivo primordial é descrito a seguir.

1.3 Objetivos

O presente estudo tem como objetivo principal fazer uso da técnica de clusterização, para realizar o agrupamento de imóveis com características semelhantes, que não seriam facilmente classificáveis pelas metodologias usualmente disponíveis para pesquisa em bancos de dados, tal como a **mineração de dados**, metodologia que consiste em utilizar técnicas automáticas de estatística e de inteligência artificial para explorar grandes massas de dados à procura de relações ou padrões úteis para as organizações, segundo Carvalho (2005) e Andrade (2009, p.22).

Consiste em identificar regras de formação dos imóveis agrupados, contribuindo para a seleção daqueles que apresentem possíveis distorções tributárias e favoreçam uma melhor compreensão dos atributos contidos no Cadastro Imobiliário. Propõe-se, com isso, a prática da justiça fiscal através das ações fiscalizadoras e das consequentes retificações cadastrais em massa, aumentando a receita tributária do município com base em uma metodologia científica e não apenas no conhecimento histórico, muitas vezes restrito a alguns servidores ou setores da Administração.

Como objetivo secundário e parte indissociável do processo, deseja-se localizar incongruências de cadastro usualmente decorrentes de falha humana na alimentação dos dados (como a digitação do valor 1,00m² para uma área que deveria ser de 100,00m²), mas que são igualmente danosas à tributação, quer por omissão na informação de dados, quer por cadastramento notadamente incompatível com o fato real apurado *in loco*, nos casos em que o sistema de controle imobiliário é incapaz de evitar tais distorções. Tal percepção decorre de pré-processamento específico que valida as características dos imóveis e indica a necessidade de correção. Ainda que identificada a incorreção no pré-processamento, os dados originais são utilizados quando da aplicação do modelo neural, para que sejam corretamente agrupados com outras distorções semelhantes.

Simple inconsistências nos bancos de dados são de elementar identificação. Por exemplo, imóveis cadastrados como possuindo dois pavimentos, mas classificados como não edificadas, ou seja, como se fossem apenas terrenos, são identificados com uma mera filtragem dos dados. Como alguns dos parâmetros da regra para tributar terrenos, são distintos daqueles para tributar imóveis edificadas, aí já se percebe uma possibilidade de perda de arrecadação decorrente da incorreção cadastral.

No modelo desenvolvido, algumas regras de inconsistência cadastral ou de suspeições geram um alerta para a adoção das verificações e retificações necessárias. São identificados, dentre outros: a) imóveis com área edificada, porém cadastrados como terreno; b) imóveis prediais, porém sem área edificada; c) imóveis públicos não cadastrados como imunes (o que evitaria incorreções na previsão da receita tributária, bem como na cobrança do IPTU enviada mensalmente pelos Correios); d) terrenos com classificação arquitetônica de casa; e) imóveis com valor venal menor que R\$ 500,00.

Em contrapartida, o agrupamento de imóveis com semelhanças em suas características, é tarefa deveras complexa e que demanda uma enorme quantidade de cálculos matemáticos, o que inviabiliza seu sucesso por mero uso de filtros ou cruzamento de dados.

1.4 Justificativas

Os cadastros técnicos imobiliários utilizados pelas secretarias de governo responsáveis pela tributação, fiscalização e arrecadação de tributos municipais, especialmente o Imposto sobre a Propriedade Predial e Territorial Urbana (IPTU), mostram-se usualmente imprecisos e vulneráveis, tanto em decorrência da dinamicidade da expansão imobiliária sem a correspondente capacidade dos fiscos de acompanhar tais evoluções, quanto em razão da ineficiência dos mecanismos de administração e controle de acesso aos dados. A imprecisão e a vulnerabilidade poderiam ser reduzidas a partir de uma remodelagem dos sistemas imobiliários, dando-lhes críticas mais específicas e consistentes.

Devido à complexidade da determinação do valor do solo em massa, base tributária que, em conjunto com as características das edificações contidas em cada parcela, subsidiam o cálculo e o lançamento do IPTU, surgem inúmeras inconsistências capazes de afetar de modo considerável o valor final do tributo que deveria ser recolhido aos cofres públicos. A ocasional omissão na padronização, normatização, instrução e verificação de dados, no que diz respeito ao Cadastro Imobiliário, aumenta a instabilidade e reduz a confiabilidade dos cadastros, especialmente pela carência de controles especializados que evitem distorções prejudiciais à ação fiscal. Observe-se, por exemplo, os casos em que não resta absolutamente esclarecida e normatizada a forma como deve ser calculada a área edificada em relação aos beirais, ou a maneira como devem ser classificadas edificações típicas como postos de gasolina. Seriam estes meros galpões, ou lojas onde é comercializada a mercadoria combustível? Há grande diferença na tributação em decorrência de uma e de outra classificação arquitetônica.

A tributação do IPTU, cuja fórmula de cálculo para o município de Fortaleza citada por Gomes e Mesquita (2003, p.161) é descrita no Anexo B, é altamente complexa, envolvendo regras matemáticas, lógicas e mesmo seletivas que definem o valor do tributo a ser pago. Devido a esta complexidade, consolidam-

se o risco, do lado do Fisco, e a oportunidade, do lado do contribuinte, de enquadrar de forma inexata as características do imóvel cadastrado, de forma a reduzir, por ação própria ou de forma omissiva, o valor a ser recolhido ao Erário. Neste sentido, a criação de minuciosa normatização que esclareça pontos polêmicos, dúbios ou divergentes, pode reduzir, juntamente com a adoção dos adequados procedimentos de correição, a evasão fiscal.

Ademais, a conexão entre bancos de dados tributários, por vezes, mostra-se ineficaz, quando permite a alteração de dados cadastrais pertinentes ao contribuinte, sem a concomitante atualização dos dados referentes ao próprio imóvel por ele ocupado, mais uma vez acarretando distorções, para menos, no valor do tributo cobrado. Verifica-se tal fato, por exemplo, quando é criada uma prestadora de serviços que passa a funcionar numa antiga residência, sem que o Cadastro Imobiliário tenha atualizado o uso específico do dito imóvel de residencial para comercial. A alíquota aplicável à base de cálculo do IPTU, conforme o caso, seria majorada em até 150%, o que, em não sendo feito, causará expressivo prejuízo para a arrecadação municipal. Uma vez mais, a carência de definição normativa para os procedimentos, permite que haja perda de receita tributária para o município, o que poderia ser corrigido com a criação de uma comissão permanente de normatização que cuidasse de discutir, propor, criar, divulgar, informar e controlar mecanismos que estimulassem a perfeita interação entre os setores da Administração, e os procedimentos por eles adotados.

1.5 Metodologia

A utilização dos bancos de dados tributários municipais, especialmente os relativos a informações de caráter imobiliário, foi condição *sine qua non* para o embasamento do trabalho proposto. Uma vez disponibilizado o acesso a tais dados, com o devido cuidado e atenção para não haver desrespeito à regra de sigilo fiscal imposta pelo artigo 198 do Código Tributário Nacional, tal como mencionam Pinto, Windt e Céspedes (2008, p.36), procedeu-se à elaboração de uma metodologia, para classificar, caracterizar e quantificar os imóveis constantes do Cadastro Imobiliário através do uso de uma rede neural.

Foram coletados, a partir do banco de dados imobiliário, campos relativos às parcelas de solo e às edificações, incluindo a situação fiscal do imóvel (tributável

ou não), a área do terreno, a área edificada, seu valor venal estimado e outros, de forma a permitir a elaboração de recomendações e ações que levassem à otimização do processo de controle da ocupação urbana através de uma política fiscal efetiva e justa.

A metodologia empregada caracteriza-se como pesquisa exploratória com estudo de caso, com as seguintes etapas:

- a) Seleção de um distrito com grande diversidade de características físicas e construtivas, representativas do conjunto de imóveis da cidade, bem como onde houvesse expressiva tributação;
- b) Elaboração de bases de dados com elementos de caracterização física e tributária dos terrenos e edificações que compõem a base imobiliária do município;
- c) Elaboração de base de dados auxiliar com os centroides das quadras onde estão localizados os imóveis, para fins de definição do posicionamento espacial dos imóveis no território do município;
- d) Construção de um ambiente inteligente sob a forma de *software*, voltado aos seguintes objetivos, dele constando serviços, funções e restrições operacionais específicas, a fim de facilitar a obtenção de informações pelo usuário:
 - Identificação e dimensionamento dos campos que contêm as características dos terrenos e das edificações, com definição dos valores possíveis de serem atribuídos a tais campos, através da criação de arquivos de apoio;
 - Análise dos dados com a concomitante segmentação dos imóveis por distrito onde aqueles foram agrupados e suas características registradas para processamento analítico dos campos envolvidos;
 - Estatísticas, com uso de medidas de posição tal como explicitado em Fonseca e Martins (1996, p. 52), através da emissão de relatórios conclusivos, com totalização de dados por distrito, bem como de acordo com as características imobiliárias que causam maior impacto tributário, quando incorretas, tais como o tipo de imóvel (territorial ou predial), a classificação arquitetônica (casa, apartamento, loja, galpão e outros), o uso específico (residencial ou não-residencial), a área do terreno e a área edificada;

- Montagem do plano cartesiano representativo do distrito que se deseja analisar, contendo a menor e a maior latitude e longitude possível para um imóvel pertencente ao distrito;
 - Visualização espacial da concentração de imóveis em determinado distrito ou no município como um todo. Para o georeferenciamento da plotagem toma-se a coordenada geográfica do centro da quadra onde está localizado cada imóvel. Podem ser visualizados os quantitativos relativos aos imóveis em geral, às áreas dos terrenos ou das edificações, às classificações arquitetônicas ou outros, havendo a diferenciação dos quantitativos por meio da plotagem de um círculo, cujo raio varia de acordo com a quantidade encontrada;
 - Criação de um ambiente gráfico que permite a rápida e fácil movimentação do usuário pelas quadras do distrito, com a concomitante visualização da concentração computada para a característica sob análise.
- e) Análise dos dados com a criação de tabelas comparativas contendo elementos que permitissem a fácil identificação de distorções tributárias. Por meio do “Relatório geral por atributo (planilha)” podem ser visualizadas totalizações, bem como os valores máximo, mínimo e médio em cada distrito, para dados como a quantidade de casas, de apartamentos, de terrenos e de imóveis em geral, dentre outros. Como exemplo de utilização das informações geradas na análise, identificou-se que o distrito com maior quantidade de apartamentos de cobertura, com 89 unidades, estaria na periferia de Fortaleza, no bairro do Barroso. O bairro do Meireles, com a maior quantidade de apartamentos da cidade, teria apenas 60 deles. Em rápida verificação identificou-se que houve o cadastro equivocado de todo um condomínio vertical no bairro do Barroso, no qual todos os apartamentos foram cadastrados como sendo de cobertura;
- f) Sugestão de atividades corretivas, a fim de otimizar a tributação e a arrecadação do IPTU no município.

No estudo de caso apresentado neste trabalho, foram aplicadas técnicas de clusterização, ou seja, separação em grupos (*clusters*), a fim de reunir em um mesmo conjunto os imóveis com características semelhantes.

1.6 Principais contribuições

As Redes Neurais Artificiais (RNA) possibilitam, tal como lembram Rao e Srinivas (2003, p.8), a execução de tarefas de maneira distribuída, em paralelo, muito mais rápida que o modo de trabalhar em série, utilizado pelos computadores convencionais. Com a aplicação desta metodologia específica para fins tributários, a Administração Pública tem um novo e expressivo aliado em sua árdua tarefa de identificar distorções tributárias causadoras de evasão fiscal. Quando se fala em arrecadação, cada centavo que entra nos cofres públicos abre novas possibilidades de investimento público em benefício da população. Imagine-se, então, um ganho de vários milhões de reais advindos da agregação da tecnologia aos procedimentos atualmente em uso. No aspecto financeiro, a contribuição que traz este trabalho, muito especialmente para a Administração Tributária municipal é tão relevante quanto o acréscimo de receita tributária causado por sua utilização.

No que tange ao aspecto social e político, a identificação e a correção de distorções causadoras de excesso, na cobrança do tributo, apresenta-se como instrumento de justiça fiscal, adequando a carga tributária à capacidade contributiva, isto é, como se depreende de Machado (2005, p.59-60), à capacidade de pagar o imposto realmente correspondente à sua propriedade. Ademais, dá ao Fisco maior credibilidade junto à população, melhorando a consciência de que se deve pagar o imposto justo devido. Incrementa, também, os mecanismos de acompanhamento, auditoria e controle com vistas a proporcionar um combate ainda mais eficaz à sonegação fiscal, cujas sanções previstas na lei de crimes contra a ordem tributária (lei 8137 de 27/11/1990) e citadas por Machado (2005, p. 480) parecem não ser suficientes para desestimular alguns contribuintes a se omitirem de recolher o valor devido, a título de imposto, aos cofres públicos.

Para a comunidade acadêmica, o trabalho também se mostra esclarecedor porquanto, ainda que não aborde em profundidade o tema das Redes Neurais, dissecar os principais conceitos a elas relacionados e traz inúmeras ilustrações que ajudam a simplificar o entendimento e a grandeza de sua aplicação.

A explicitação passo-a-passo de todo o aprendizado, revelada através do algoritmo de treinamento, do fluxograma de implementação da rede e do relatório de ajuste de pesos, para os interessados no tema, facilita a implementação de RNA semelhante para uso em aplicativo específico, acadêmico ou profissional.

1.7 Organização do trabalho

Buscou-se, com o trabalho, apresentar não apenas as Redes Neurais Artificiais, mas especialmente os conceitos necessários à sua fundamentação e entendimento, trazendo sempre ilustrações que valessem por mil palavras. Todas as figuras apresentadas foram criadas pelo próprio mestrando.

O assunto foi introduzido no primeiro capítulo, que também traz uma abordagem sobre os problemas identificados, as hipóteses, os objetivos, as justificativas, a metodologia e as principais contribuições trazidas com a realização dos estudos e com a aplicação efetiva das técnicas de clusterização.

Como a tecnologia das RNA foi aplicada e implementada em um estudo de caso englobando matéria tributária, a dissertação contém, em seu segundo capítulo, vasta abordagem sobre como acontece a tributação e os elementos que a permeiam, a fim de facilitar o entendimento da aplicação prática, discorrendo-se também sobre a estrutura de um cadastro imobiliário, suas características e limitações.

No terceiro capítulo, é apresentada toda a fundamentação teórica julgada oportuna e necessária, a fim de esclarecer os principais conceitos sobre as Redes Neurais Artificiais. São tecidas considerações sobre a estrutura cerebral e a relação dos neurônios biológicos com os neurônios artificiais empregados nas RNA. O mecanismo, através do qual os neurônios adquirem um novo conhecimento, é ampla e detalhadamente apresentado, quando são explicados tanto o treinamento supervisionado como o não supervisionado, amplamente comentados por Haykin (2001, p.75-137). Algumas fórmulas matemáticas, entendidas como úteis, especialmente para aqueles que desejem implementar *softwares*, também foram comentadas. Todavia, tentou-se torná-las tão claras quanto possível, a fim de que a matemática que elas trazem em seu bojo não se tornassem elemento dificultador do entendimento para os que não são tão afeitos a este ramo do conhecimento.

Noções relevantes para o bom entendimento do funcionamento da rede também foram detalhadas, tal como o procedimento para normalização de dados, que, segundo Priddy e Keller (2005, p. 15-20), acelera o processamento e reduz as distorções entre os dados. As medidas de similaridade, a definição e utilização dos critérios de parada e a determinação da vizinhança também foram amplamente abordados. Por fim, no mesmo capítulo, a fundamentação teórica é concluída com a apresentação da Rede Neural de Kohonen, também chamada de Mapas Auto-

organizáveis de Kohonen, ou *Self-Organizing Maps (SOM)*, em que não existe resposta previamente definida para a entrada especificada, segundo Beale e Jackson (1990, p.110). As redes de Kohonen são especialização maior das Redes Neurais Artificiais, e tecnologia aplicada no caso prático, exposto nesta dissertação,

No quarto capítulo, são sugeridos procedimentos de preparo de um cadastro imobiliário com a efetiva aplicação das RNA, com estudo de caso do município de Fortaleza, sendo descrito o equipamento utilizado e o software desenvolvido exclusivamente pelo mestrando para a demonstração da metodologia de classificação apresentada na presente dissertação.

No quinto capítulo são analisados trabalhos correlatos que trazem a aplicação das Redes Neurais a temas de cunho imobiliário. No sexto capítulo, as conclusões são reportadas de forma analítica e consistente, no sentido de justificar a realização de todo o estudo descrito e implementado nos demais capítulos.

Em virtude da grande quantidade de termos que podem ser desconhecidos para muitos leitores, decidiu-se, também, utilizar um glossário antecedendo os anexos como forma de facilitar a referência e a explicação dos novos termos apresentados.

2 CONCEITOS TRIBUTÁRIOS

2.1 Considerações iniciais

Segundo a Constituição Federal de 1988, comentada em Pinto, Windt e Céspedes (2008, p.9), o município é uma das unidades que compõem a República Federativa do Brasil, autorizando-o a instituir impostos dentre os quais o Imposto Sobre a Propriedade Predial e Territorial Urbana (IPTU), cuja competência tributária é atribuída à municipalidade no inciso I do artigo 156 da Carta Magna. Ficam, assim, sob a égide do município, não só a responsabilidade de tributar, mas também de fiscalizar e de arrecadar o imposto em questão, através do órgão responsável pelas finanças municipais.

O Código Tributário Nacional, que, entre outros elementos, define a natureza jurídica¹ específica do tributo, como explica Alexandre (2008, p.41), e é também comentado por Pinto, Windt e Céspedes (2008, p.7), lei nº 5172 de 25 de outubro de 1966 (hoje com *status* de lei complementar), define, na cabeça de seu art. 32, que são fatos geradores² do IPTU a propriedade, o domínio útil ou a posse de bem imóvel localizado na zona urbana do Município. A ocorrência de tais fatos faz incidir anualmente sobre cada parcela de solo, quer edificado ou não, um crédito tributário em desfavor do contribuinte, assim entendido o proprietário ou o possuidor do imóvel, valendo esclarecer, como bem ensina Machado (2005), que o crédito tributário é um vínculo jurídico, por força do qual, o Estado pode exigir do particular o pagamento do tributo ou da penalidade pecuniária.

Considerando que a tarefa de manter o equilíbrio entre receitas e despesas municipais passa pela busca incessante de otimizar a arrecadação tributária e que, no caso específico do IPTU, a arrecadação tem estreita ligação com a capacidade do Fisco de identificar corretamente o imóvel do contribuinte e as características que influenciam no cálculo do imposto a ser recolhido aos cofres municipais, o Erário tem especial interesse financeiro na retenção e atualização de dados, quantitativos e qualitativos, da malha imobiliária municipal.

¹ A natureza jurídica do tributo é determinada pelo fato gerador da obrigação tributária.

² Fato gerador é aquela situação que, quando acontece, gera a obrigação de pagamento do imposto. Ensina Alexandre (2008) que o fato gerador da obrigação tributária principal (ou seja da obrigação de pagar) é a situação definida em lei como necessária e suficiente à sua ocorrência, sendo portanto situação jurídica.

2.2 Elementos da tributação imobiliária

Para efeito de melhor entendimento, e de forma simplificada, cabe esclarecer que o IPTU é calculado a partir da estimativa do valor venal³ do imóvel, tributariamente chamado de **base de cálculo**. Sobre este valor é aplicada uma **alíquota**, ou seja, um determinado percentual conforme cada caso específico, resultando no imposto a pagar.

O valor venal do imóvel tem em sua correta aferição uma tarefa complexa a se processar, ainda que dependa unicamente de dois elementos: o valor do terreno e o valor da edificação.

O valor do terreno, mesmo decorrendo basicamente de sua localização geográfica e de suas características físicas (imóveis de esquina, por exemplo, são, em geral, mais valorizados que os demais), mostra-se demasiadamente dinâmico, uma vez que é influenciado pela existência ou ausência de benfeitorias públicas ou empreendimentos privados nas proximidades, tanto quanto por limitações impostas em decorrência da intervenção do Poder Público em função do zoneamento urbano e da desapropriação de imóveis para a construção de equipamentos públicos. A cada dia, a implantação de novas praças, escolas, supermercados, hospitais, favelas, postos de gasolina, delegacias de polícia, *shopping centers*, casas de espetáculos, igrejas e outros, exercem particular influência, positiva ou negativa, na valorização da parcela de solo.

Especialmente em virtude da já mencionada dinâmica de valorização do solo, torna-se impraticável, ineficiente e demasiadamente onerosa a aferição individualizada do valor de cada parcela de solo no território do município. Desta forma, no sentido de viabilizar uma avaliação em massa dos valores dos terrenos, é criada a chamada **Planta Genérica de Valores Imobiliários** (PGVI), que especifica o valor do metro quadrado dos terrenos de acordo com sua localização e tem, como uma de suas vantagens, a redistribuição da carga fiscal, permitida através da redução da diferença entre o valor fiscal e o valor de mercado, conforme Möller (1995 apud BRONDINO, 1999, p.25). É comum que os valores sejam os mesmos para todos os imóveis pertencentes a uma mesma face de uma mesma quadra.

³ Valor venal do imóvel é o valor que o imóvel alcançaria no mercado imobiliário no momento de uma venda à vista, em condições normais.

O valor da edificação depende da área construída no terreno, de sua idade, de seu estado de conservação, do tipo de material utilizado no acabamento da obra, de sua classificação arquitetônica (apartamento, casa, loja, galpão...) dentre outros fatores.

Uso do imóvel	Especificações	Alíquota
Territorial	Área sem infra-estrutura (pavimentação, iluminação pública e água)	1,0%
	Área com infra-estrutura, com muro e calçada	1,6%
	Área com infra-estrutura, sem muro e sem calçada	2,0%
Residencial	Até R\$ 58.500,00	0,6%
	De R\$ 50.500,01 até R\$ 210.600,00	0,8%
	Acima de R\$ 210.600,00	1,4%
Não residencial	Até R\$ 210.600,00	1,0%
	Acima de R\$ 210.600,00	2,0%

Quadro 1 – Variação na alíquota do IPTU do município de Fortaleza.

Já a alíquota tem um valor objetivo e de fácil aferição a partir de regras bem definidas. Imóveis residenciais, por exemplo, possuem alíquotas específicas de acordo com seu valor. Já os não residenciais (comerciais, industriais, prestadores de serviço...), possuem outra. Os imóveis territoriais têm, ainda, uma alíquota distinta dos residenciais e dos não residenciais. Basta, então, identificar o uso específico do imóvel, i.e. se ele é residencial, não-residencial ou territorial, para a correta aplicação da alíquota a partir de seu valor venal, conforme cada caso. O Quadro 1, com base em Gomes e Mesquita (2003), ilustra a variação da alíquota do IPTU no município de Fortaleza, conforme o uso do imóvel e as especificações que indica, podendo variar de acordo com a legislação própria de cada município.

Uma vez que a abordagem tributária não é o objetivo principal deste trabalho, a fim de simplificar o entendimento, omitiu-se o comentário sobre um elemento conhecido como **reductor**, cujo objetivo é garantir a justiça fiscal através da correta aplicação da alíquota conforme as faixas de valores venais. O conhecimento do mecanismo proporcionado pelo reductor é, no entanto, de leitura mandatória para a inteira análise dos efeitos tributários.

Percebe-se, então, que a tributação resta dificultada pela tarefa complexa de determinar, com necessária precisão, as características de cada unidade imobiliária. A incorreta caracterização usualmente prejudica a eficiência fiscal e compromete seriamente tanto a fiscalização como, especialmente, a arrecadação do imposto. Mais que isso, fragiliza a relação de justiça fiscal entre o ente público e o contribuinte, ocasionalmente aferindo uma base de cálculo maior que a, de fato, aplicável, fazendo pesar sobre aquele que tem o dever de pagar o tributo, a árdua missão de provar que existe um excesso tributário causado pela incorreta aferição das características do imóvel, no caso, o foco da matéria tributável.

Por outro lado, convém frisar, que é dever do contribuinte comunicar ao Fisco a alteração de características que afetem a tributação de seu imóvel. Por exemplo, não é raro acontecer que o contribuinte edifica área adicional e se furta de comunicar o ente responsável pela tributação, muitas vezes no anseio de evitar qualquer majoração no imposto a ser pago.

2.3 O cadastro imobiliário

O Cadastro Imobiliário é a representação da estrutura física sobre a qual agirá o Fisco na sua complexa tarefa de tributar. Entende-se por cadastro o depósito de dados referentes a uma determinada matéria. Quando o cadastro tem como finalidade o armazenamento de dados relativos a imóveis, é então chamado de Cadastro Imobiliário. Nele são encontrados diversos tipos de dados com o objetivo de armazenar suas características e demais elementos a eles relacionados. Certamente, é interesse da Administração Tributária fazer constar no Cadastro Imobiliário todos os dados indispensáveis à correta tributação e tantos quantos mais sejam necessários, a fim de evitar e coibir a evasão fiscal, isto é, a redução ou a omissão deliberadas do pagamento do tributo devido.

2.3.1 Características do cadastro

Como a tributação imobiliária é complexa, pelos motivos já expostos anteriormente, criar e manter um banco de dados de grandes proporções tende a ser tarefa complexa e onerosa. A cada cadastro de imóvel criado é atribuído um número de inscrição. Em um prédio residencial onde existem 100 apartamentos pertencentes

a pessoas distintas, existirão 100 inscrições, sendo uma para cada apartamento. A cada inscrição ficarão ligados os dados do imóvel para efeito de atualização, pesquisa, relatórios, cruzamentos de informações e outros.

Devem ser guardados os dados relativos à área do imóvel, à sua data de construção, se o imóvel é de esquina, o tipo de material usado no acabamento de pisos, paredes e teto, a existência de equipamentos de lazer, a classe arquitetônica (casa, apartamento, galpão, sala...) e muitas outras características que permitem avaliar com razoável grau de certeza o valor do imóvel.

No Anexo B podem ser encontrados alguns exemplos de campos utilizados para composição de uma base de dados imobiliária, ou para o cálculo do imposto propriamente dito, tais como:

- a) Fator de depreciação – percentual utilizado para reduzir o valor da edificação em função do desgaste natural causado pela sua idade;
- b) Fração ideal – proporção do terreno que cabe ao proprietário de uma edificação erguida sobre um lote onde existe mais de uma edificação, como no caso dos apartamentos, por exemplo;
- c) Fator de correção do lote – valor calculado para valorizar ou desvalorizar o lote de terreno em função de suas características (normal, esquina, gleba e outros) ou da existência de melhorias urbanas (água, esgoto, iluminação, pavimentação e outros).

Vale lembrar que, para efeito do Cadastro Imobiliário e, conseqüentemente, deste estudo, o termo **imóvel** representa cada **inscrição imobiliária**. Desta forma, um condomínio vertical residencial com 20 apartamentos, possui 20 imóveis cadastrados.

Imagine-se, então, um banco de dados imobiliário que contenha 500 mil inscrições cadastradas. Se cada inscrição estiver ligada a pelo menos 30 dados diferentes, ter-se-ão armazenados 500.000 vezes 30, ou seja, 15 milhões de dados, isto sem considerar outros tantos relativos ao proprietário do imóvel ou ao pagamento do imposto ao longo dos anos.

2.3.2 Atualização cadastral

Ainda que haja a obrigação legal do contribuinte de informar quaisquer alterações capazes de alterar a tributação, o desconhecimento da norma e até

mesmo a falta de medidas de coerção eficazes, findam por colaborar, para que o cadastro não permaneça atualizado ao longo do tempo.

O imóvel nasce para o cadastro no momento de sua inclusão, assim entendida a primeira vez em que os dados relativos ao imóvel são inseridos na base de dados imobiliários da Administração. Durante a inclusão é de suma importância que os dados alimentados sejam criticados, manual ou eletronicamente, passando por testes de consistência no sentido de que não sejam registrados dados que não traduzem a situação real do imóvel. A identificação de inconsistências, através do cruzamento de informações, é mandatória nos casos em que o cadastro foi criado e vem sendo mantido sem a devida crítica durante a alimentação de dados.

Para ilustrar a relevância da validação dos dados, dá-se o seguinte exemplo: como já vimos, a Planta Genérica de Valores Imobiliários (PGVI) define os valores do metro quadrado do terreno para cada face de quadra do município. Nos casos de terrenos com mais de uma face, algumas legislações consideram, para fins de cálculo do IPTU, que o valor definido para o metro quadrado do terreno é aquele da face de maior valor, tal como acontece no município de Fortaleza. Desta forma, nos casos de imóveis sujeitos à tributação, se para um determinado imóvel de esquina existe no cadastro apenas a face de menor valor cadastrada, está ocorrendo clara evasão fiscal. Há casos em que, após cadastrada a face secundária de maior valor, o terreno tem seu valor quadruplicado. Nos casos em que existe apenas uma face de quadra cadastrada, quando o tipo do lote (esquina, quadra ou gleba) sugere que deveria haver mais de uma, pode estar ocorrendo perda de receita tributária, caso a face omissa seja a de maior valor. Cada ocorrência precisa ser identificada de maneira particular, quer por meio de fotografias aéreas, com verificações *in loco*, ou por outro meio, para posterior retificação.

Outro exemplo de inconsistência cadastral passível de solução através do cruzamento de informações acontece, quando o Cadastro Imobiliário indica que um determinado imóvel é utilizado para fins exclusivamente residenciais, quando, de fato, existe empresa constituída e operante no endereço do imóvel em questão. Vale lembrar que as alíquotas aplicáveis a imóveis não-residenciais usualmente são bem mais elevadas do que as aplicáveis a imóveis residenciais, mais uma vez caracterizando perda de receita tributária para o Erário.

Em ambos os casos, a correção dos problemas passa necessariamente pela atualização cadastral, fazendo constar os dados corretos sobre os imóveis e

restando ao Fisco a cobrança do imposto pago a menor, apenas para o período ainda permitido pela legislação tributária. Como o instituto da decadência descrito no artigo 173 do Código Tributário Nacional (lei 5.172 de 25/10/1966) traz, em linhas gerais, que somente podem ser lançados⁴ impostos no prazo retroativo de 5 anos além daquele em curso, todo o imposto anterior a este período não mais poderá ser cobrado, constituindo muitas vezes expressiva lesão aos cofres públicos.

Com isso, percebe-se que a desatualização cadastral, em geral, importa cobrança a menor do imposto devido, motivo pelo qual manter o cadastro atualizado tem consumido elevado volume de recursos e despertado sobremaneira a atenção dos Fiscos. Não bastasse considerar a evolução na valorização imobiliária decorrente de atividades especulativas por parte de grandes e pequenos latifundiários urbanos, a dinâmica no financiamento de moradias, bem como os investimentos do governo e de particulares na construção civil, quer em grandes empreendimentos, quer em pequenas reformas, contribuem sobremaneira para tal desatualização.

Adicionalmente, a extração de informações a partir dos dados cadastrados é de suma importância para o planejamento tributário realizado de forma continuada pela Administração, aí incluídas ações fiscais e de tributação tendentes a alavancar a arrecadação. O que dizer, então, se tais ações forem tomadas com base em dados desatualizados e que não refletem a realidade?

Quando relatórios de inconsistências ou mesmo cruzamentos de informações fiscais não se mostram suficientemente eficazes no sentido de perceber, no universo dos dados, a existência de distorções cadastrais, a agregação de metodologia científica, que extraia informações não facilmente perceptíveis, demonstra ter grande utilidade, especialmente em virtude da complexidade da obtenção de tal informação.

Uma vez percebido o potencial que tem uma desatualização cadastral no sentido de distorcer as características do imóvel e, conseqüentemente, de subestimar a sua tributação, traduzida no valor do imposto a ser recolhido pelo contribuinte aos cofres públicos, multiplique-se este potencial pela quantidade de

⁴ O lançamento tributário é a atividade que compete à autoridade administrativa e que implica na constituição do crédito tributário a ser pago pelo contribuinte. Vide artigo 142 *caput* do Código Tributário Nacional (lei 5172/66).

imóveis desatualizados e ter-se-á a ordem de grandeza do prejuízo causado pela desatualização aos cofres públicos.

2.3.3 Riscos nas avaliações em massa

As consequências naturais da desatualização cadastral são o cálculo e a cobrança inexatos do imposto. A enorme quantidade de imóveis que normalmente compõe os cadastros imobiliários municipais inviabiliza sua atualização pontual, imóvel por imóvel, forçando a uma atualização em massa muito especialmente quando do reajuste da Planta Genérica de Valores Imobiliários (PGVI) ou quando da aplicação de qualquer outro índice que tencione recompor o valor de um grupo determinado de imóveis, como, por exemplo, os imóveis não-residenciais.

Na verdade, em virtude das administrações públicas disporem de recursos humanos e materiais insuficientes e, mais ainda, em decorrência da escassez de mecanismos seguros que permitam obter - no caso do IPTU - uma aproximação ótima da estimativa feita para o valor venal dos imóveis, estes findam por ser subavaliados, redundando em uma tributação aquém daquela que poderia ser efetivamente aplicada.

Para evitar que as avaliações imobiliárias ultrapassem o valor de mercado dos imóveis, causando descontentamento generalizado dos contribuintes atingidos pela avaliação excedente, é comum que os Fiscos optem, como medida preventiva, por subavaliar, com largo nível de segurança, os imóveis como um todo, reconhecendo, com isso, a incapacidade de otimizar, em níveis elevados, a aferição dos valores venais e, conseqüentemente, da arrecadação.

2.3.4 Distribuição espacial

A localização espacial do imóvel é deveras relevante. Primeiro, porque tem influência em seu valor e, conseqüentemente na tributação e na arrecadação. Exemplo disso é que os terrenos situados na orla marítima - via de regra - têm melhor avaliação que aqueles situados na periferia das cidades. Segundo, porque, a partir de sua localização, podem ser identificadas inúmeras distorções cadastrais. Seria razoável supor, a título de exemplo, uma valorização expressiva dos imóveis

localizados nas redondezas de onde foi construído o primeiro grande empreendimento comercial de um bairro, como um *shopping center*, por exemplo.

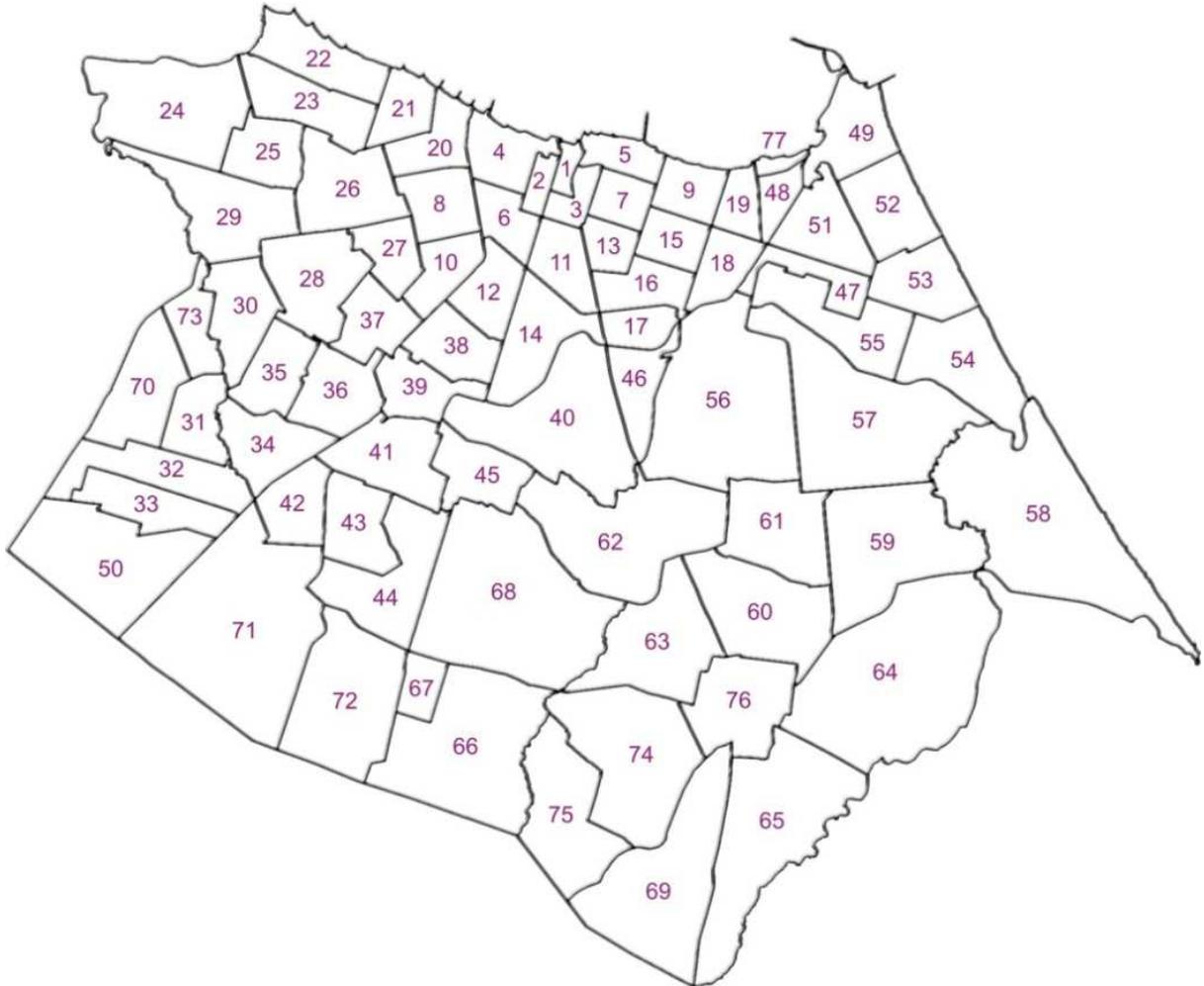


Figura 2 – Divisão do município de Fortaleza em distritos

Outro exemplo que ilustra a importância da identificação de distorções cadastrais advém da observação no cadastro imobiliário, de mais de 500 imóveis para os quais não constava haver pavimentação asfáltica para veículos, em uma avenida pela qual transitam centenas de milhares de veículos diariamente. A omissão cadastral desta melhoria urbana, em geral, é fonte de perda de receita tributária, enquanto o adequado cadastramento do asfalto na pavimentação para veículos, nos termos da legislação tributária do município de Fortaleza, usualmente aumenta o valor venal do imóvel, acrescentando o valor do IPTU devido.

O município de Fortaleza se encontra dividido em 77 distritos, tal como ilustrado na Figura 2. Já a Figura 3 ilustra como um determinado distrito pode estar

subdividido em quadras. No caso, um dos distritos que compõem o Centro da cidade de Fortaleza.

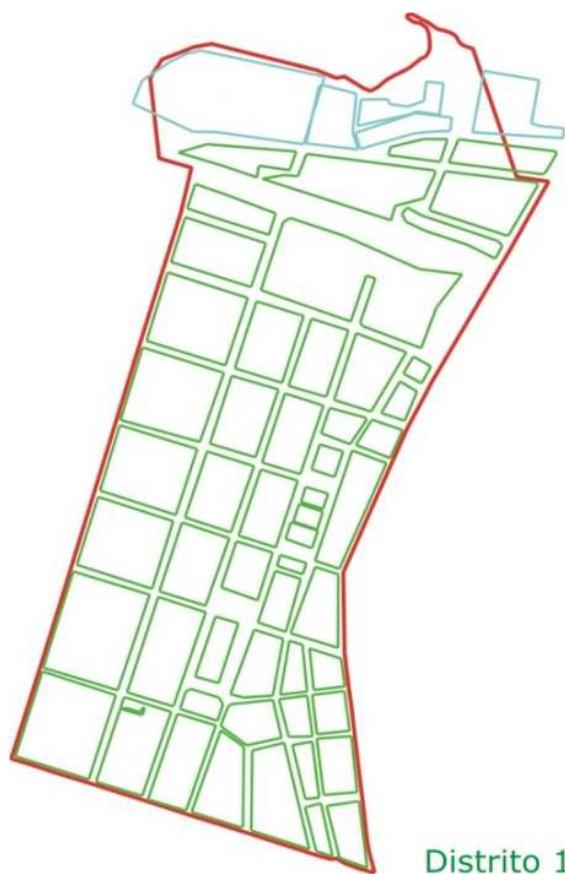


Figura 3 – Subdivisão do distrito 1 (parte do bairro Centro) e suas quadras.

Apesar das tecnologias já disponíveis para localizar, através de coordenadas georeferenciadas, um ponto qualquer sobre o território nacional, alguns cadastros imobiliários ainda têm como base de localização a **cartografia** do imóvel, o que equivale à sua coordenada na forma descrita no Quadro 2, ou seja, distrito, quadra, lote e complemento do lote.

O **distrito** é criado de acordo com a lei que também define o polígono que o delimita. Cada **quadra** é criada a partir do nascimento de loteamentos específicos de grandes parcelas de terra, que subdividem o terreno em partes menores com o fim de exploração comercial. É daí que nasce cada **lote** de terra, ou seja, a menor parcela divisível do terreno. Em um mesmo lote, quando construídas diversas unidades imobiliárias, diferencia-se umas das outras através de um número denominado de **complemento do lote**. É o caso de condomínios residenciais verticais, quando em um mesmo lote de terra é construído um prédio com inúmeros

apartamentos, recebendo cada um deles um número sequenciado de complemento, pois todos se encontram edificados sobre a mesma parcela de solo.

Como a localização do imóvel é um dos fatores que compõem a valorização e a tributação imobiliárias, a correta indicação da cartografia, ou o referenciamento geográfico, conforme o caso, são imprescindíveis para os Fiscos municipais. Desta forma, um eventual equívoco no cadastramento do distrito, da quadra ou mesmo do lote relativo ao imóvel, pode comprometer seriamente a tributação, dando causa, inclusive, à evasão fiscal.

Distrito	Quadra	Lote	Complemento
77	14	121	3

Quadro 2 – Exemplo de cartografia tradicional.

2.4 Considerações finais

É possível concluir que a tributação imobiliária é complexa e delicada, uma vez que se encontra permeada por inúmeras variáveis e circunstâncias cuja equivocada valoração fragiliza a correta aferição da matéria tributável. O valor venal real dos imóveis é, portanto, de difícil determinação. Estimativas comumente imprecisas, levam o Fisco a subavaliar o imposto a pagar uma vez que a desatualização cadastral ao longo dos anos tende a crescer até a adoção de um sempre oneroso, porém recompensador, recadastramento em massa.

É perceptível, também, que somente uma atualização cadastral continuada minimiza os efeitos negativos da incoerência entre os dados cadastrais e os dados reais que refletem a situação dos imóveis *in loco*. Neste ponto, a própria Administração precisa perceber, o quanto é vital a contratação de atualizações periódicas em massa, em benefício dela própria e, porque não dizer, mesmo, da população em favor da qual a arrecadação de impostos retorna através de obras públicas nas mais diversas áreas, como saúde, educação, moradia e outras.

A utilização de imagens aéreas, obtidas de satélites ou voos autorizados, traduz-se em boa fonte indicativa de distorções de áreas edificadas, sendo possível adquiri-las com boa relação custo/benefício. No entanto, para um completo e eficaz cadastro, é necessária a realização de visita ao imóvel, a fim de levantar as

características que irão permitir maior exatidão na apuração do valor venal, base de cálculo do IPTU a pagar.

Por fim, a Administração Tributária carece de mecanismos científicos que, além da metodologia tradicional de exploração de dados, permitam ir além da lógica booleana, do tipo SE → ENTÃO, para extrair relações entre os dados imobiliários, viabilizando a correção do cadastro através de ações específicas e facilitando a identificação de distorções que geram evasão fiscal para os cofres públicos.

Ao serem mencionadas ações não-lógicas, pretende-se estender o entendimento de que um novo formato de busca de informações precisa ser implantado. Um modelo de raciocínio em paralelo, muitas vezes mais eficaz que o modelo serial e lógico do computador convencional, mas que, com a ajuda dele, possa ser simulado. Um modelo metodológico baseado no funcionamento dos neurônios cerebrais, a fonte do raciocínio humano, o núcleo de todo o aprendizado, é a seguir exposto amiúde como proposta para aumento de eficácia da Máquina Fiscal e instrumento de justiça tributária.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 Considerações iniciais

Seriam os computadores mais inteligentes que os humanos? Calcular a soma de todos os números primos entre um e dez mil pode ser uma tarefa trivial para um computador, mas certamente não o será para humanos. No entanto, a identificação de um objeto ou a movimentação de um braço para pegá-lo, tarefa que o cérebro executa com inquestionável destreza, é imensamente difícil para um computador. Segundo Beale e Jackson (1990), os computadores são máquinas lógicas e só conseguem fazer bem coisas lógicas.

O computador funciona como uma estrutura montada sobre milhões de minúsculos circuitos e componentes eletrônicos conectados entre si de forma ordenada, onde as operações são conduzidas de forma seriada, ou seja, uma após a outra. Com enorme capacidade de processamento e memorização, o computador é capaz de substituir a presença humana de modo sensivelmente eficiente nas tarefas onde há demanda expressiva por cálculos, comparações e memorização.

Já o cérebro em nada nos lembra a estrutura retilínea e organizada de uma placa de circuitos de um computador. Nele, é possível encontrar uma complexa estrutura com bilhões de microscópicos elementos que, como os computadores, também tratam impulsos elétricos, só que de forma simultânea, ou seja, em paralelo: os neurônios.

Na verdade, estruturas tão distintas como o computador e o cérebro não poderiam trabalhar da mesma forma.

3.2 A estrutura cerebral

O neurônio é a unidade básica do cérebro, sendo auto-suficiente em termos de capacidade de processamento. De acordo com Beale e Jackson (1990) os neurônios podem ser de dois tipos: a) células de processamento local entre neurônios com conexões de entrada e de saída; e b) células de saída que conectam diferentes regiões do cérebro, ou estas, com músculos e outros órgãos do corpo.

Quando uma determinada quantidade de solicitações é feita ao neurônio, ele é ativado, saindo do seu estado de repouso. O corpo da célula, do neurônio, é chamado de **Soma**. A ele se encontram ligados filamentos irregulares chamados de **dendritos**, através dos quais chegam as solicitações ao neurônio, ou seja, suas entradas. Outro componente do neurônio é o **axônio**, elemento eletricamente ativo, ao contrário do dendrito, e que funciona como um canal de saída do neurônio. Ao pulso elétrico produzido pelo axônio dá-se o nome de **potencial de ativação**. Não é comum encontrar axônios nas conexões entre aqueles neurônios em que tanto entradas quanto saídas ocorrem a partir dos dendritos. O axônio termina no ponto de conexão com o dendrito de outra célula, a **sinapse**. Tal conexão é temporária e acontece através dos **neurotransmissores**, elementos químicos liberados pela sinapse, quando o potencial de ativação alcança níveis apropriados. Algumas sinapses excitam os dendritos, outras os inibem, o que equivale a alterar a direção do fluxo.

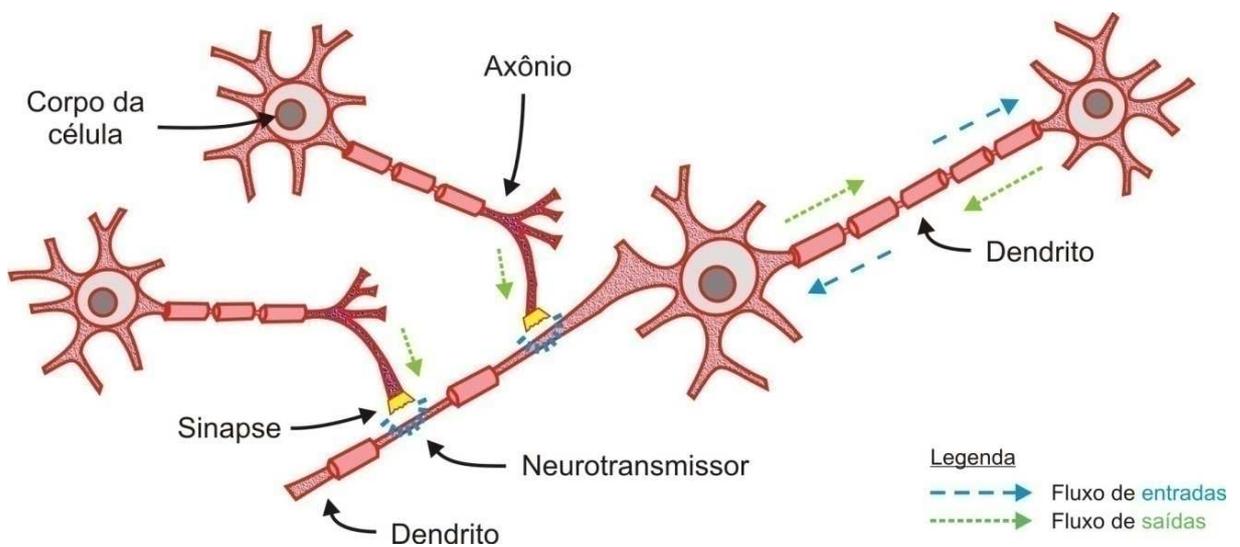


Figura 4 – Estrutura elementar de um neurônio cerebral

Em resumo, temos que os neurônios cerebrais, tal como ilustrado na Figura 4, são compostos pelo corpo da célula, pelos dendritos e pelo axônio, comunicando-se por meio de conexões sinápticas ativadas quimicamente pelos neurotransmissores. Os dendritos recebem informações de outros neurônios que são conduzidas ao corpo da célula, onde ocorre o processamento, sendo a informação posteriormente encaminhada, através do axônio, até os dendritos de outros neurônios.

Apesar da condução dos impulsos elétricos estabelecida na estrutura cerebral ser muito mais lenta do que a realizada pelos computadores, o fato do processamento cerebral acontecer em paralelo, ou seja, com a execução de várias tarefas de uma só vez, dá-lhe expressiva vantagem, quando se trata de realizar muitas tarefas simultaneamente. Assim, segundo Beale e Jackson (1990), quando existe necessidade de se processar grande número de informações relacionadas entre si, na busca de uma solução desejada, com o conhecimento vindo de diferentes fontes, cada qual dando sua própria contribuição, o essencial não é a velocidade e, sim, o paralelismo proporcionado pelo cérebro, cuja maior capacidade demonstrada é a de aprender coisas de forma simples, rápida e objetiva, e de a elas se adaptar. Por outro lado, apenas através de extensos, complicados e comumente imperfeitos programas, os computadores ousam simular a capacidade cerebral de paralelismo.

No cérebro, um determinado trabalho é dividido entre vários neurônios, ao invés de deixar todo o trabalho para uma única estrutura de processamento. Este modelo de trabalho traz algumas vantagens, dentre elas o fato de minimizar a importância da participação de cada neurônio no processo como um todo. Logo, se um dos neurônios falha ou morre, a tarefa não fica comprometida.

A distribuição do trabalho, da forma que é feita pelos neurônios cerebrais, é conhecida como **processamento distribuído**, o que qualifica a estrutura cerebral a se auto-ajustar, quando ocorrer perda permanente de um neurônio, substituindo-o por um outro. A habilidade de prosseguir com o funcionamento da tarefa, mesmo depois da perda do neurônio, sem que seja gerada uma solução inconsistente ou vazia para o problema tratado, é chamada de **tolerância a falhas**.

Adicionalmente, em casos de danos permanentes, o sistema paralelo distribuído reduz vagarosamente sua performance de um nível alto para um nível baixo, no entanto sem interromper drasticamente o processamento. Esta redução paulatina possibilita ainda a transferência de parte ou de todo o processamento afetado para outras estruturas, ou mesmo a obtenção de soluções parciais aproximadas para o problema apresentado. Este procedimento é conhecido como **degradação graciosa**, conforme Beale e Jackson (1990), e se traduz em uma vantagem adicional em favor do cérebro, quando comparado com a capacidade limitada dos computadores de auto-reconstruir um processo prejudicado.

3.3 Redes Neurais Artificiais (RNA)

A fim de construir um modelo em que os computadores conseguissem agir tanto quanto possível como o cérebro humano, em 1943 o neurofisiologista Warren McCulloch e o matemático Walter Pitts, da Universidade de Illinois, nos Estados Unidos, desenvolveram as Redes Neurais Artificiais (RNA) através de um sofisticado trabalho que fazia analogia entre as células nervosas vivas e o processamento eletrônico binário utilizado pelos computadores. Foi descrito o modelo artificial de um neurônio e foram apresentadas suas capacidades computacionais. Segundo Braga, Carvalho e Ludemir (2000), a base de uma RNA é a estrutura individual dos nodos, as conexões e o comportamento dos neurônios naturais, que processam e se comunicam com milhares de outros, continuamente e em paralelo no cérebro humano, estrutura que contém 100 bilhões deles.

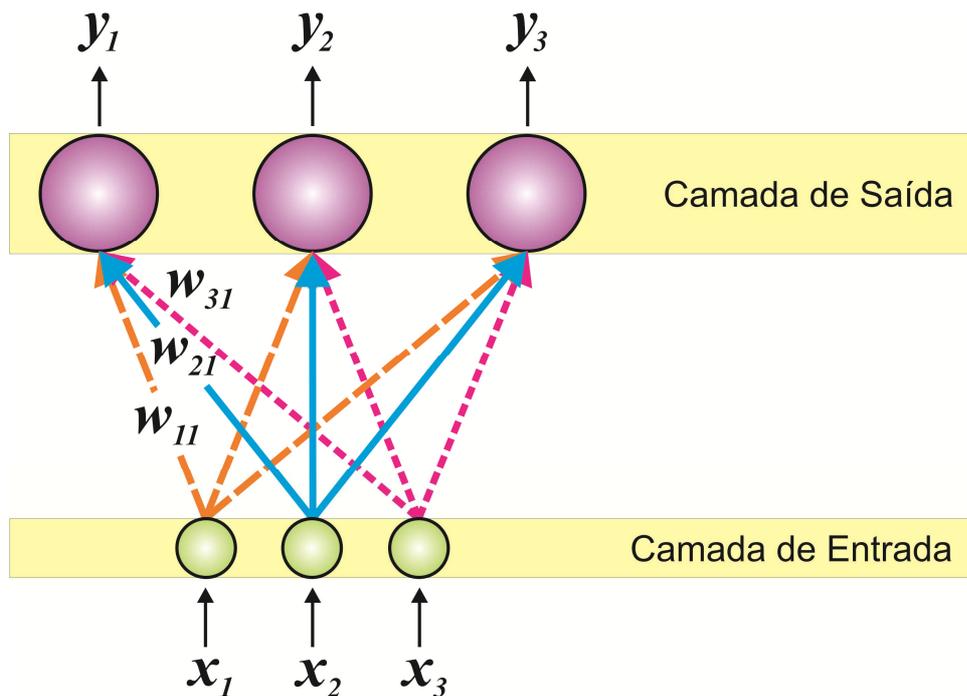


Figura 5 – Rede com associações lineares em duas camadas

Em uma Rede Neural Artificial, existem associações lineares com pares de vetores (x_i, y_i) . Quando o vetor x_i é apresentado como entrada à rede, o vetor de saída correspondente deve ser y_i . Fazer com que a rede responda adequadamente a cada nova entrada é, portanto, tarefa a ser alcançada pela correta valoração da matriz de pesos (w_{ij}) entre as ligações x e y , tal como exemplificado na Figura 5, adaptada de Rao e Srinivas (2003), onde:

x_i → entrada i apresentada à rede;

y_j → saída j construída pela rede;

w_{ij} → peso da entrada i em relação à saída j no tempo t ;

Logo, uma RNA é formada basicamente por uma Camada de Entrada e uma Camada de Saída, onde se encontram posicionados os respectivos neurônios. Também compõem a rede pesos sinápticos (w_{ij}) que indicam a força da conexão entre os neurônios da Camada de Entrada (x_i) e os neurônios da Camada de Saída (y_i).

A **Camada de Entrada** apresenta à rede neurônios aleatórios a serem analisados (x_1, x_2, \dots, x_n) e classificados em grupos semelhantes. Para cada nova entrada a rede trabalhará no sentido de agregar novo conhecimento, ou seja, novo aprendizado ao já existente, ajustando-se em relação às ligações entre seus neurônios.

A **Camada de Saída** contém um número específico de neurônios representativos de cada agrupamento a ser formado e disposto em topologia apropriada. Antes do efetivo processamento da rede, deve ser definida a quantidade de neurônios que irão formar esta camada, ou seja, a quantidade máxima de agrupamentos finais propostos. No caso da Figura 6, a estrutura desta camada conta com nove neurônios, identificados como y_1, y_2, \dots, y_9 , que darão origem a até nove agrupamentos distintos na medida em que as entradas forem sendo apresentadas à rede.

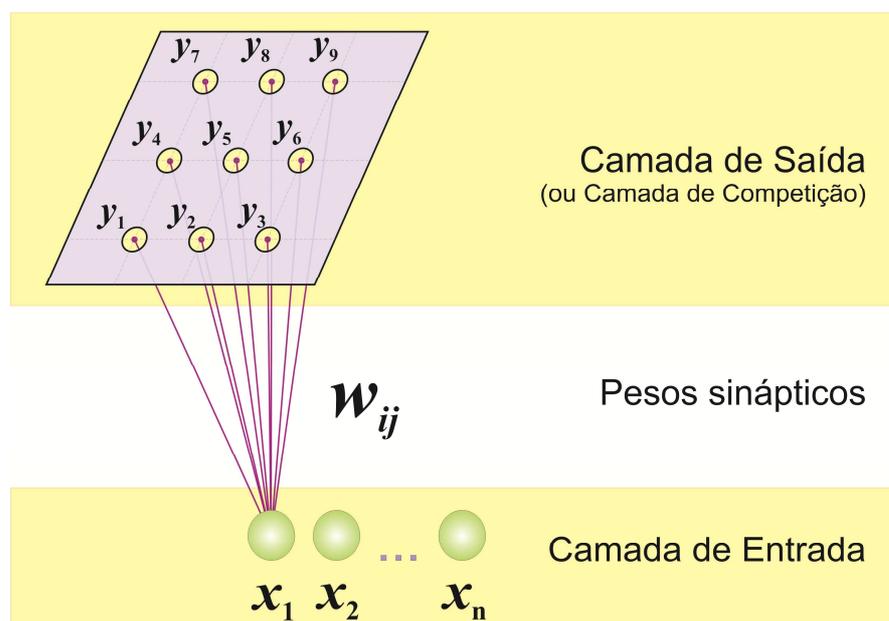


Figura 6 – Estrutura de relacionamento entre neurônios

O vetor de **pesos** (w_{ij}) representa a força de conexão entre a Camada de Entrada e a Camada de Saída, e é inicializado aleatoriamente com valores bem próximos de zero, definindo que existe, *a priori*, uma fraca ligação entre os neurônios das Camadas de Entrada e de Saída. Os pesos serão adequadamente ajustados ao longo do processamento da rede. Frise-se que o conhecimento adquirido pela rede ao longo do processamento reside justamente no peso de cada conexão, destacam Rao e Srinivas (2003). Normalmente tanto as entradas como os pesos são normalizados, o que significa que a magnitude de ambos é ajustada para 1.

Em linhas gerais, o objetivo da RNA é organizar as entradas em neurônios de saída tão similares quanto possível, sendo que cada neurônio representará um grupo com características semelhantes. Naturalmente, quando o Operador da Rede define uma maior quantidade de neurônios, na Camada de Saída, está permitindo que o processamento classifique as entradas com um maior nível de detalhamento.

3.4 Aprendizado

As maiores partes do cérebro são os hemisférios, que são dispostos em camadas e cuja porção mais externa equivale à parte mais complexa da estrutura, o **córtex cerebral**. Os neurônios que o formam apresentam elevada interconectividade, em virtude da alta densidade de suas células. Experimentos realizados com animais demonstraram que sua remoção implica a perda da capacidade de aprender e de se adaptar, frisam Beale e Jackson (1990), levando à dedução de que o córtex é o núcleo da inteligência e das funções mais refinadas do cérebro. Ainda, de forma experimental, restou comprovado que o desenvolvimento organizado do córtex cerebral não está única e exclusivamente ligado a fatores genéticos, mas também ao ambiente e a seus eventos, quando o córtex exprime sua habilidade de aprender e de se auto-organizar a partir de estímulos externos.

De forma análoga, a aprendizagem conseguida com os computadores tem sido implementada através da observação e percepção de informações capturadas a partir do ambiente externo. Assim, a solução de um problema para o qual o conhecimento é inicialmente limitado passa por um processo de aprendizagem. A este processo, dá-se o nome de **aprendizado** ou **treinamento**,

quando são identificados padrões de entradas e procede-se aos seus ajustes interativos, através da atribuição de pesos, de acordo com a relevância da informação obtida, ensinam Braga, Carvalho e Ludemir (2000). A propositura de um algoritmo de aprendizado deve ser acompanhada de um modelo que exprima a teoria por ele proposta, quando podem ser utilizadas expressões algébricas no sentido de modelar a adaptação dos pesos relativos a cada informação de entrada, lembram Rao e Srinivas (2003).

Em 1949, no livro "*The organization of behaviour*", o psicólogo canadense Donald Hebb descreveu um mecanismo razoável para o aprendizado a nível celular, no cérebro, difundindo a ideia de que, quando uma entrada no neurônio causa a imediata emissão de um pulso, então existe expressiva habilidade desta entrada, para gerar futuramente um pulso no neurônio de destino, ocorrendo aumento na eficácia da sinapse que transmite o sinal de entrada para o neurônio de saída, frisam Rao e Srinivas (2003). Ou seja, a conexão sináptica tem aumentada a sua relevância, que é representada pelo seu peso. De outra forma, se a entrada não causa qualquer reação no neurônio de destino, a conexão sináptica tem diminuído o seu peso, a sua importância, mostrando baixa ou nenhuma relação entre os neurônios de entrada e de saída.

Os algoritmos de aprendizado são formados basicamente por uma **regra de aprendizado**. Sua base biológica foi inicialmente proposta por Donald Hebb, em 1949. Uma regra de aprendizado pode ser de dois tipos principais: o primeiro, chamado de *regra de Hebb* é utilizado, quando apenas a entrada é conhecida; o segundo, chamado de regra *Delta* ou regra dos **erros mínimos quadrados**, é utilizado quando tanto a entrada quanto a saída desejada são conhecidas. Adaptações destas regras, com modificações simples, geram inúmeras outras regras em uso na atualidade.

O aprendizado das redes neurais normalmente é iterativo, significando que o mesmo procedimento é repetido para um certo número de entradas. A estas iterações, dá-se o nome de **ciclos**. Após cada ciclo, as entradas e os pesos sinápticos podem ser alterados ou permanecerem constantes. Denomina-se **convergência** a propriedade que as regras de aprendizado devem possuir de não corromper ou invalidar informações anteriormente obtidas, mas de usá-las no sentido de aproximar o resultado de uma solução ótima, assim entendida aquela que

representa uma perfeita similaridade entre os dados classificados em um mesmo agrupamento.

Quando, após um determinado ciclo, não há mais alterações nas entradas ou nos pesos, indicando que houve convergência dos valores e que não mais existe aprendizado, é desnecessário prosseguir com ele. Diz-se, neste caso, que foi alcançada a **estabilidade** na rede. Em muitos casos a convergência excede o tempo razoável proposto para o aprendizado da rede, devendo ser parametrizado um nível de tolerância aceitável, a fim de que, mesmo sem ter sido alcançada plena convergência, a rede seja considerada treinada e o aprendizado seja encerrado de forma precoce, ainda que satisfatória, segundo Rao e Srinivas (2003).

Em resumo, tem-se que o objetivo do aprendizado é inferir a saída ótima desejada, através do mapeamento das entradas, o que é feito através de comparações entre as entradas e as saídas, sendo estas últimas previamente definidas ou aleatoriamente inicializadas. Existem dois tipos principais de aprendizado: o **supervisionado** e o **não-supervisionado**.

3.4.1 Aprendizado supervisionado

No aprendizado supervisionado, o processo percebe o estímulo do ambiente externo, ou seja, as entradas, e conhece as saídas desejadas. Portanto, cada par <entrada>/<saída desejável> é previamente conhecido, cabendo ao processo ajustar os pesos das entradas, a fim de minimizar o erro no caminho que leva à saída desejada. Segundo Braga, Carvalho e Ludemir (2000), para um dado tempo t , o erro $e(t)$ pode ser calculado pela Equação 1:

$$e(t) = d(t) - y(t)$$

Equação 1 – Erro $e(t)$ entre a saída calculada e a saída desejada

sendo:

$d(t)$ → peso desejado no tempo t ;

$y(t)$ → peso conforme a resposta calculada pelo algoritmo de aprendizado no tempo t .

A Equação 2 mostra a forma genérica para ajuste dos pesos através da minimização da diferença entre a soma ponderada dos pesos das entradas e a saída desejada, considerando-se o respectivo erro especificado.

$$\mathbf{w}_i(\mathbf{t} + 1) = \mathbf{w}_i(\mathbf{t}) + \boldsymbol{\eta}\mathbf{e}(\mathbf{t})\mathbf{x}_i(\mathbf{t})$$

Equação 2 – Ajuste de pesos entre as conexões

sendo:

$\mathbf{w}_i(\mathbf{t} + 1)$ → peso da entrada i no tempo $t+1$;

$\mathbf{w}_i(\mathbf{t})$ → peso da entrada i no tempo t ;

$\boldsymbol{\eta}$ → taxa de aprendizagem entre 0 e 1;

$\mathbf{e}(\mathbf{t})$ → erro no tempo t ;

$\mathbf{x}_i(\mathbf{t})$ → entrada i no tempo t .

Entendendo, como *professor*, um supervisor externo que fornece os pares <entrada>/<saída desejada>, e, como *rede*, a estrutura composta por todas as entradas e saídas, Braga, Carvalho e Ludemir (2000) apresentam, como desvantagem do aprendizado supervisionado, o fato de que na ausência do professor, a rede não conseguirá aprender novas estratégias para situações não cobertas pelos exemplos de pares apresentados. No plano da eficácia, significa dizer que a rede não seria capaz de inferir, por si só, a associação entre a entrada e a saída desejada.

Para que fosse usada esta técnica de aprendizado, no caso específico da classificação de imóveis, seria necessário criar vasto número de correlações entre os diversos tipos de imóveis existentes, e os grupos aos quais deveriam ser associados, fornecendo, portanto, conhecimento prévio à rede. Assim, somente após a associação preliminar mencionada, seria possível a classificação de outros dados diversos daqueles já classificados pelo Operador da Rede.

3.4.2 Aprendizado não-supervisionado

Em virtude da forma como informações são armazenadas, o cérebro humano esquece com relativa facilidade aquilo que conheceu ou aprendeu, diferente dos computadores. O cérebro armazena informações de maneira distribuída, descentralizada, ao contrário dos computadores que as armazenam de forma

centralizada. Seres humanos lembram com mais intensidade daquilo que viveram nos últimos dois minutos do que daquilo que viveram há duas décadas, isto, porque a recuperação de informações antigas pelo cérebro é consequência da adaptação de conexões sinápticas entre neurônios, tanto quanto da **memória associativa**, ou seja, da capacidade de associação entre a informação a ser recuperada e outras informações a ela relacionadas direta ou indiretamente.

No aprendizado não-supervisionado não são conhecidas saídas desejadas. Ao contrário, a rede, ao receber uma entrada, percebe o estímulo do ambiente externo e, através de um processo de auto-organização, agrupa informações de naturezas semelhantes. Acontece então um auto-aprendizado através do ajuste dos pesos das entradas em relação às saídas, a fim de permitir que uma nova entrada apresentada à rede seja corretamente associada ao neurônio (grupo) adequado da Camada de Saída a partir de sua natureza, de suas características.

São importantes para a rede, tanto o tempo gasto, para que ela seja considerada treinada, quanto o tempo de classificação de uma nova informação entre os grupos formados. É desejável que exista uma **tolerância** em relação a distorções na classificação de novas entradas, mas o algoritmo de aprendizado deve buscar o erro mínimo aceitável, o que deve ser especificado, como parâmetro de aprendizado da rede.

No aprendizado não-supervisionado apenas as entradas são conhecidas. A quantidade de neurônios da Camada de Saída, também conhecida como **Camada de Competição** neste tipo de aprendizado, é definida pelo Operador da Rede. Os neurônios da Camada de Saída competirão entre si, para que seja identificado qual deles possui maior similaridade com a entrada apresentada.

Os pesos entre cada par <entrada/saída> são aleatoriamente inicializados com valores próximos de zero, a fim de que sejam ajustados com o decorrer do treinamento. O ajuste de pesos é realizado pela Equação 3 descrita a seguir:

$$w_{ij}(t+1) = \begin{cases} w_{ij}(t) + \alpha(t) \times (x_i(t) - w_{ij}(t)), \forall j \in \Lambda(t) \\ w_{ij}(t), \forall j \notin \Lambda(t) \end{cases}$$

Equação 3 – Ajuste de peso do neurônio vencedor e de seus vizinhos

onde :

$w_{ij}(t + 1)$ → peso entre o neurônio de entrada i e o neurônio j da Camada de Competição no instante $t+1$;

$w_{ij}(t)$ → peso entre o neurônio de entrada i e o neurônio j da Camada de Competição no instante t ;

$\alpha(t)$ → taxa de aprendizagem no instante t ;

$x_i(t)$ → neurônio de entrada i no instante t ;

$\Lambda(t)$ → raio de vizinhança do neurônio vencedor no instante t ;

Vale frisar a fundamental importância da taxa de aprendizagem alfa (α), onde $\alpha \in (0,1)$, através da qual o ajuste dos pesos é procedido. A taxa que é inicializada com valores altos entre 0 e 1, é decrementada, com o decorrer do tempo, até atingir valores bem pequenos, próximos a zero, quando a aprendizagem se torna altamente especializada. Quando a taxa de aprendizagem possui valores altos, próximos de 1, a rede rapidamente esquece velhos aprendizados, mantendo apenas os mais recentes e fazendo com que a matriz de pesos cresça sem a adequada limitação, ensinam Rao e Srinivas (2003).

Diante dos conceitos apresentados, convém definir **clusterizar** como selecionar grupos semelhantes de elementos. Tais elementos são apresentados como entrada e, a partir da análise de similaridade entre suas características, faz-se sua correta classificação.

3.5 Normalização de dados

Antes de iniciar o treinamento da rede, todas as entradas são normalizadas, i.e. transformadas para uma mesma escala, na qual, passam a ser medidas, ensina Späth (1980, p.21). No presente estudo todas as variáveis foram normalizadas para valores entre 0 e 1.

Um dos recursos mais úteis na otimização de sistemas de reconhecimento é a normalização de dados, que equivale a fazer com que cada uma das características dos dados esteja em um intervalo único pré-definido, proporcionando, assim, uma distorção mínima no processamento da rede neural, quando duas características forem comparadas entre si, ou seja, as variáveis são

transformadas para uma mesma escala na qual passam a ser medidas, como ensina Späth (1980, p.21). Esta pesquisa apresenta todas as variáveis transformadas para valores entre 0 e 1.

Para que ocorra a normalização, o Operador da Rede deve decidir previamente em quantos intervalos distintos ela será processada. Os intervalos corresponderão a classes, e todos os dados dentro dos limites de cada classe serão tratados como se fossem exatamente iguais. Se a opção for normalizar utilizando-se quatro classes, então a rede terá habilidade para distinguir tipos de dados cujos valores assumam, no máximo, quatro valores distintos. Se a rede for classificar dados que podem assumir mais de quatro valores distintos, como é o caso da classificação arquitetônica (barraco, casa, apartamento, apartamento de cobertura, loja, sobreloja, sala, galpão aberto, galpão fechado e outros), duas ou mais classificações arquitetônicas distintas serão posicionadas em uma mesma classe. Com isso, a intervenção e atenção do Operador da Rede podem ser fundamentais para que se alcance um treinamento adequado, evitando a indesejável valoração equitativa de dados diferentes, tal como se pertencessem a uma mesma classe.

A normalização também se presta à aceleração do treinamento da rede, uma vez que as características passam a se enquadrar em uma mesma faixa de grandeza. Os métodos de normalização são muitos, cada um com propriedades e benefícios particulares. Quando as entradas e saídas da rede neural precisam ser restringidas a limites específicos, a aplicação de uma fórmula de interpolação linear pode ser procedida, a fim de colocar os velhos valores em uma nova escala, usualmente na faixa entre 0 e 1, inclusive.

O método Min-Máx, sugerido por Priddy e Keller (2005) e descrito na Equação 4, implementa tal interpolação e tem a vantagem de preservar exatamente as mesmas proporções dos valores originais, tal como pode ser visualizado na Figura 7. Vale destacar que, se na amostra todos os valores de entrada são constantes, eles devem ser desprezados, pois não contribuem para o aprendizado da rede neural.

$$x'_i = (max_{novo} - min_{novo}) \times \left[\frac{(x_i - min_{antigo})}{(max_{antigo} - min_{antigo})} \right] + min_{novo}$$

Equação 4 – Interpolação linear pelo método Máx-Min

onde:

x_i → valor antigo;

x'_i → valor novo, ou seja, valor antigo transformado para a nova escala;

min_{antigo} → menor valor da escala antiga;

max_{antigo} → maior valor da escala antiga;

min_{novo} → menor valor da nova escala;

max_{novo} → maior valor da nova escala;

sendo $(max_{antigo} - min_{antigo}) \neq 0$

A normalização de dados traz ainda uma peculiaridade ilustrada na Figura 7. Para variáveis ou dados discretos, i.e. aqueles que só podem assumir valores inteiros, como ensina Góes (1998), a normalização demanda tratamento especial. No entanto, em todo caso, o Operador da Rede deve parametrizar a quantidade de subdivisões nas quais os dados devem ser alocados para fins de normalização. Estas subdivisões são chamadas de **classes** e a definição de seus limites é feita com base no que ensina a estatística básica para a construção de intervalos de classes, tanto em relação às mencionadas variáveis discretas, como em relação às variáveis contínuas.

Segundo Silva *et al.* (1999), o intervalo de classe é qualquer subdivisão da amplitude total de uma série estatística. Desta forma, em relação aos dados contínuos, tem-se que:

$$A_t = X_{máx} - X_{mín}$$

Equação 5 – Amplitude total de uma amostra

$$h = \frac{A_t}{C}$$

Equação 6 – Amplitude do intervalo de classes

$$l_c = (C - 1) \times h$$

Equação 7 – Limite inferior da classe c

$$L_c = C \times h$$

Equação 8 – Limite superior da classe c

sendo, para as Equações de 5 a 8:

- $X_{mín}$ → menor valor da amostra;
- $X_{máx}$ → maior valor da amostra;
- A_t → amplitude total;
- C → número desejado de classes;
- h → amplitude do intervalo de classes;
- l_c → limite inferior da classe c ;
- L_c → limite superior da classe c ;
- c → classe com elementos no intervalo $l_c \rightarrow L_c$.

Vale frisar que, como o intervalo de classes é aberto no limite inferior (l_c), convencionou-se que o elemento equivalente a este limite na primeira classe, a ela pertenceria. O exemplo ilustrado na Figura 7 apresenta três valores contínuos, quais sejam 5, 20 e 45, normalizados como 0,1, 0,4 e 0,9 em uma situação na qual o Operador da Rede fez a opção de segmentar a amostra em 10 classes, subdivididas na forma da Tabela 1, conforme os limites explicitados na Equação 7 e ainda na Equação 8.

Em relação às variáveis discretas, tal como na Tabela 2 e também na Figura 7, por exemplo, a subdivisão em classes respeita a simples codificação original, quando usualmente são atribuídos códigos numéricos a valores não numéricos. Os códigos, já variáveis discretas, são normalizados de forma análoga àquela que acontece com as variáveis contínuas.

Tabela 1 – Subdivisão em classes dos dados contínuos de uma amostra ($l_c \rightarrow L_c$)

Classe	Limite inferior da classe (l_c)	Limite superior da classe (L_c)	Valor normalizado
1	0	5	0,1
2	5	10	0,2
3	10	15	0,3
4	15	20	0,4
5	20	25	0,5
6	25	30	0,6
7	30	35	0,7
8	35	40	0,8
9	40	45	0,9
10	45	50	1,0

Tabela 2 – Subdivisão em classes dos dados discretos de uma amostra

Classe	Código	Conteúdo associado	Valor normalizado
1	1	Casa	0,25
2	2	Loja	0,50
3	3	Galpão	0,75
4	4	Apartamento	1,00

A Tabela 2 exemplifica a segmentação das variáveis discretas de uma amostra tal como ilustrado na Figura 7, quando o Operador da Rede decidiu por subdividir os dados de acordo com os valores originais de cada neurônio. Com isso o neurônio *classificação arquitetônica*, por exemplo, restou classificado pelos quatro códigos originais encontrados, quais sejam: 1 representando o conteúdo *casa*, 2 representando *loja*, 3 representando *galpão* e 4 representando *apartamento*.

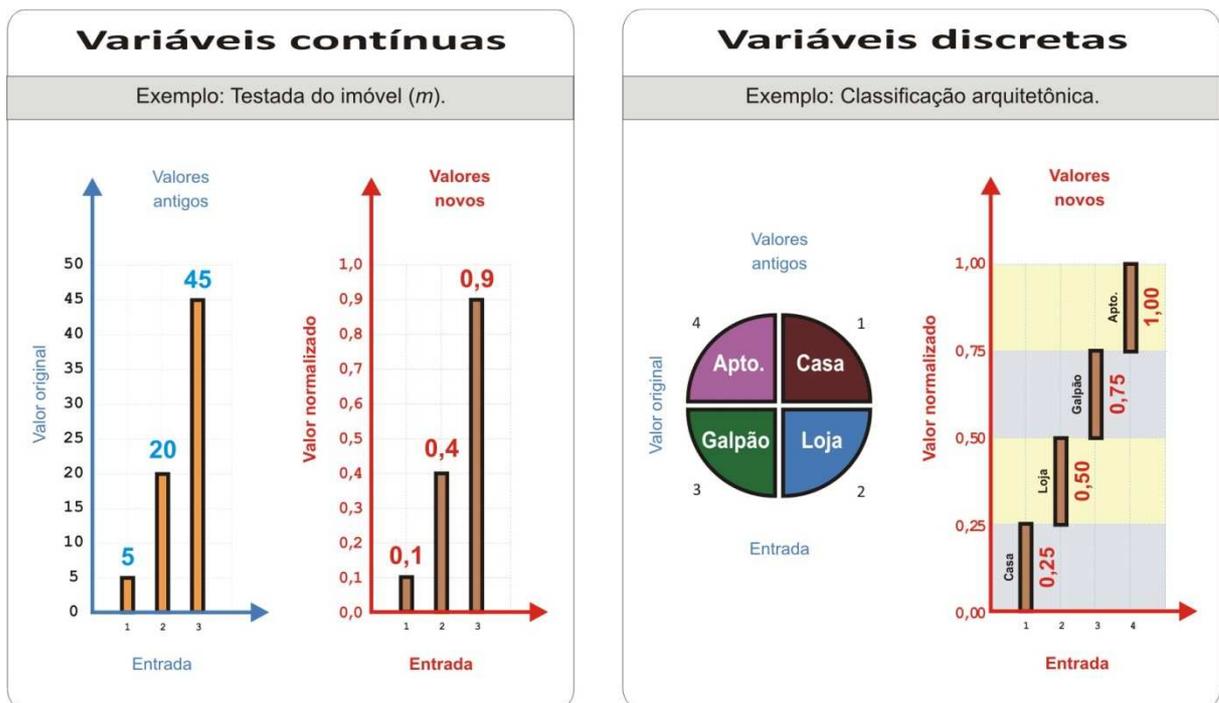


Figura 7 – Exemplo de normalização de dados para variáveis contínuas e discretas

3.6 Características dos dados

Uma determinada entrada é formada por um conjunto único de características em um espaço multidimensional, onde cada dimensão representa uma característica particular, segundo Priddy e Keller (2005). Por exemplo, um cadastro de imóveis no qual constem dados sobre a **área do terreno**, a **área**

edificada e a **classificação arquitetônica** (casa, apartamento, sala, galpão...) além de outros dados, tal como exemplificado na Tabela 3, leva a uma distribuição tridimensional das características.

Tabela 3 – Exemplo de dimensões (características) em um banco de dados imobiliário

Área do terreno (m ²)	Área edificada (m ²)	Classificação arquitetônica
400	350	Casa
200	300	Casa
1200	800	Galpão
800	50	Apartamento
1100	25	Sala
1000	950	Galpão
350	600	Casa
900	150	Apartamento
1300	60	Sala
1200	90	Sala
700	80	Apartamento
1100	700	Galpão

Percebe-se, então, que para cada entrada é necessário um vetor de características de N dimensões que ajude a especificar o imóvel, sendo tais vetores representados por pontos no espaço dimensional. Evidentemente, em um cadastro imobiliário, facilmente o espaço dimensional chegaria a possuir mais de uma dezena de dimensões.

Na abordagem analítico-explicativa proposta como exemplificação na tabela que se segue, e ainda para efeito de simplificação, foram plotadas apenas a área do terreno e a área edificada, mantendo o desenho bi-dimensional na forma apresentada na Figura 8, ao invés de serem trabalhadas dezenas de características em um ambiente multidimensional de grandes proporções, como certamente seria desejável em uma aplicação profissional.

Convém ressaltar que a seleção criteriosa das características é vital para o sucesso da Rede Neural Artificial. Quando são selecionadas poucas características e existe uma grande massa de dados, o Operador da Rede Neural pode deixá-la mensurar a importância de cada característica ao longo do processamento. Todavia, se há uma grande variedade de características e uma limitada quantidade de dados, o treinamento pode vir a produzir resultados inexpressivos em virtude de não conseguir mapear de forma clara os dados

apresentados, restando realizar o treinamento considerando apenas as variáveis mais importantes ou utilizar dados adicionais, ainda que posteriormente estes venham a ser expurgados dos agrupamentos finais.

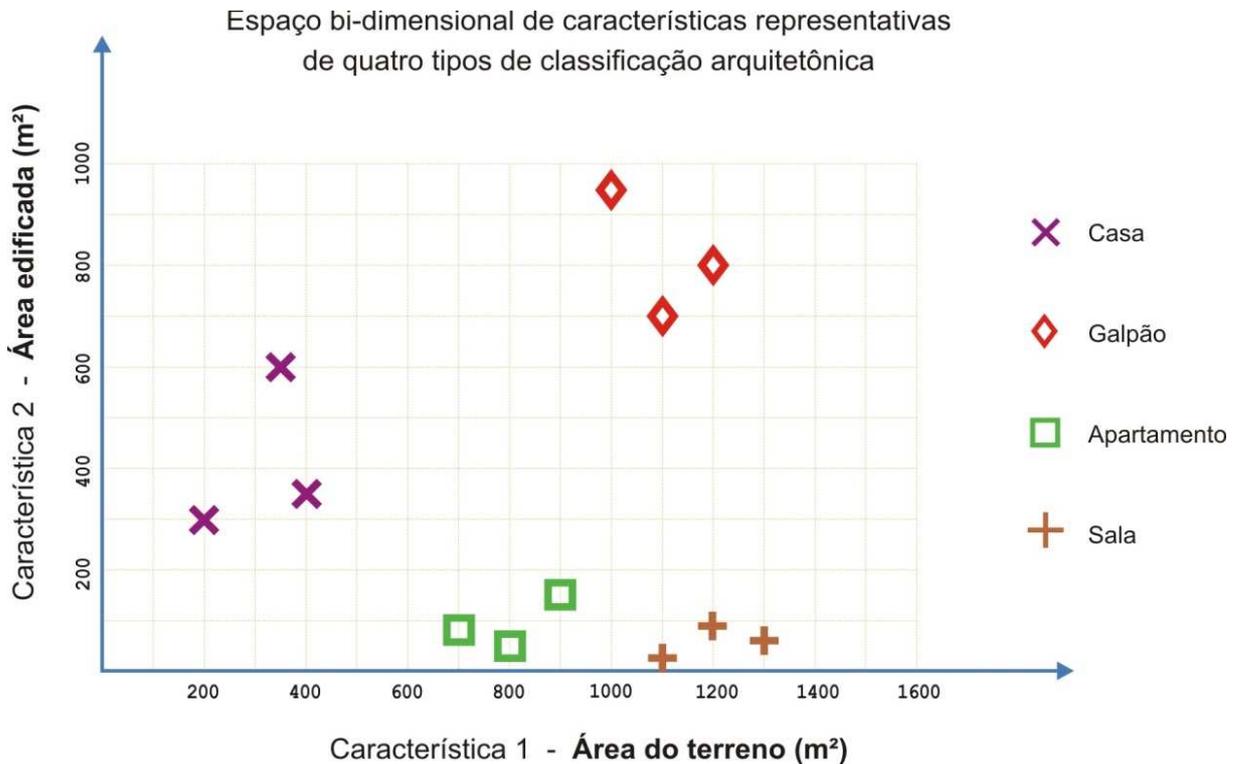


Figura 8 – Representação espacial das características do banco de dados

O processamento com um número excessivo de características também é indesejável, exigindo que as mesmas sejam limitadas, ou seja, que uma seleção de características seja feita pelo Operador da Rede, a fim de escolher quais características devem ser consideradas e quais devem ser relevadas. A parametrização da RNA deve ocorrer a partir da identificação dos tipos de dados correspondentes a cada característica, daí porque é desejável que o Operador tenha razoável familiaridade com os dados a serem analisados.

3.7 Medida de similaridade

Pode-se dizer que quase todas as técnicas de agrupamento envolvem processos de medição, quer de distâncias entre dois objetos, quer de similaridade entre eles. É o que assevera Späth (1980, p.15). A similaridade entre dois vetores de

características pode ser levantada através de várias métricas de distanciamento $d(x,y)$ em qualquer espaço com N dimensões. No presente trabalho, foi utilizada uma variação da **distância euclidiana**. Esta decorre da aplicação básica do Teorema de Pitágoras, mostrando-se bastante eficiente e de fácil visualização, tal como ilustrado na Figura 9, com base em Beale e Jackson (1990).

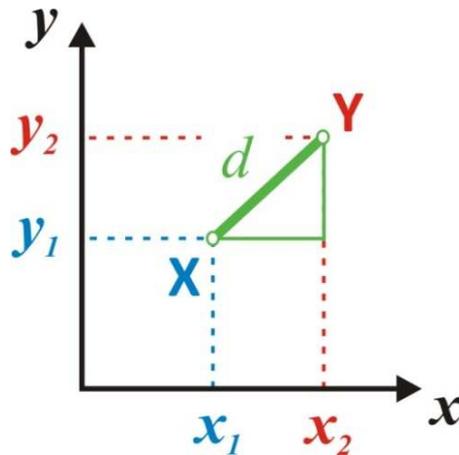


Figura 9 – Visualização da Distância Euclidiana (d) em plano bi-dimensional

No sentido de tornar mais inteligível a explicação passo-a-passo sobre o treinamento da Rede Neural, optou-se por trabalhar com o quadrado da distância euclidiana, eliminando-se, portanto, a raiz quadrada do termo apresentado na Equação 9.

Através da análise do nível de similaridade, as conexões entre as entradas e os neurônios da Camada de Saída são avaliadas, a fim de identificar o neurônio com menor distorção, ou seja, com maior proximidade de características em relação à entrada apresentada, o que é matematicamente apurado através da distância euclidiana, de acordo com a Equação 9 a seguir:

$$d(X, Y) = \sqrt{\sum_{i=1}^N (X_i - Y_i)^2}$$

Equação 9 – Distância euclidiana entre dois vetores X e Y

onde:

$d(X, Y)$ → distância euclidiana entre os vetores da Camada de Entrada X e da Camada de Saída Y;

i → índice das entradas;

- X_i → neurônio i da Camada de Entrada;
 Y_i → neurônio i da Camada de Saída;
 N → número de entradas;

Desta forma, o neurônio da Camada de Saída com maior similaridade, i.e. com maior proximidade, em relação à entrada apresentada, é chamado de **neurônio vencedor** e seu peso é incrementado, procedendo-se assim ao **ajuste de pesos** e fortalecendo sua conexão com a entrada em questão. Tendo sido identificado o neurônio vencedor, seus vizinhos são igualmente beneficiados e têm também fortalecidos os pesos de suas conexões.

3.8 Vizinhaça

A vizinhaça trata da relação de importância que têm os neurônios próximos ao neurônio vencedor. Quando uma determinada entrada é apresentada à rede e os neurônios da Camada de Saída competem entre si para identificar qual deles possui maior similaridade com a entrada, apenas um se consagra vencedor. A partir daí a rede ajusta os pesos deste neurônio em relação à entrada, ou seja, em relação a todos os neurônios que compõem a Camada de Entrada.

Topologias de vizinhaça

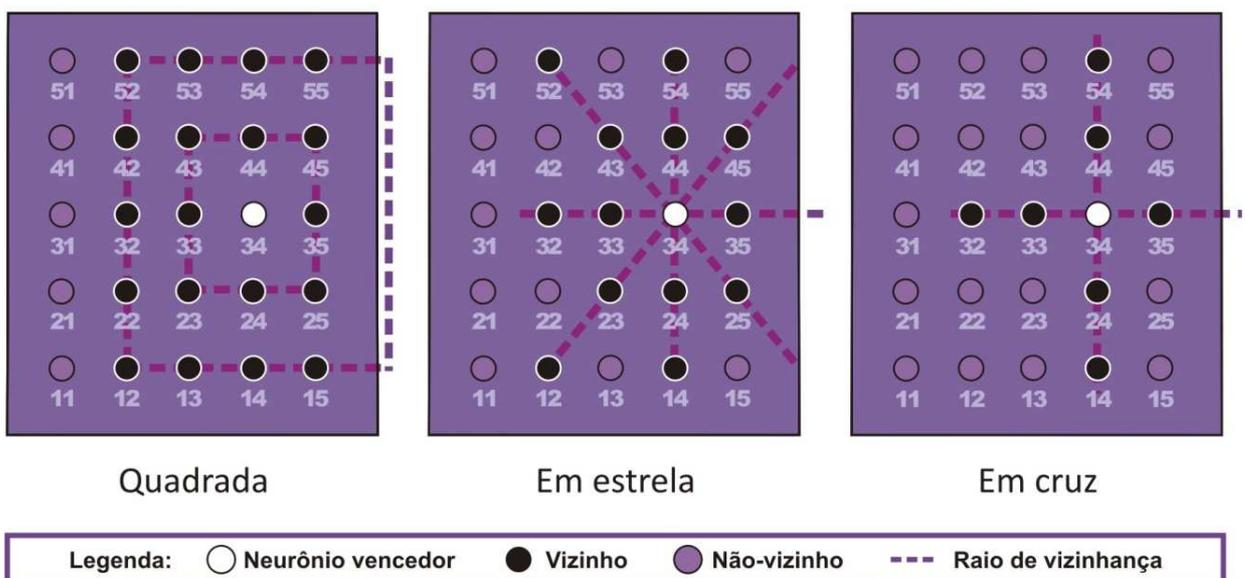


Figura 10 – Exemplos de topologias de vizinhaça do neurônio vencedor

Supondo que para uma determinada RNA, toda entrada apresentada à rede possui como características a **área do terreno**, a **área edificada** e a **classificação arquitetônica**, então estas corresponderão aos neurônios da Camada de Entrada. Se, após competirem, os neurônios da Camada de Saída têm como vencedor, para uma determinada entrada E apresentada, o neurônio localizado na linha 3, coluna 4, ou seja, na posição 34 (leia-se “três quatro”), então o peso deste neurônio será ajustado em relação à **área do terreno**, à **área edificada** e à **classificação arquitetônica** para aquela entrada.

Por sua vez, caso o raio de atuação da vizinhança tenha sido definido como de 2 unidades, então, quando um determinado neurônio vencer, os demais que estejam à distância de até dois neurônios do vencedor terão seus pesos igualmente ajustados. Mas esta distância será medida em todas as direções? Apenas nas direções ortogonais ao neurônio vencedor, ou seja, para cima, para baixo, à esquerda e à direita, respeitando as linhas e colunas nas quais estão dispostos os neurônios? Ou também poderá ser medida de forma oblíqua? O alcance da vizinhança é medido pelo seu raio, mas as direções projetadas são indicadas pela **topologia da vizinhança**.

A topologia da vizinhança pode ser quadrada, em cruz, em estrela ou pode possuir algum formato que se julgue apropriado definir conforme cada caso prático. Cabe esclarecer que a topologia é permanente durante todo o aprendizado. Já seu raio deve decrescer na medida em que o treinamento evolui e a rede se torna mais especializada com o tempo.

A topologia da vizinhança é importante por definir a velocidade de especialização que se deseja para o aprendizado. Quanto mais neurônios influenciados pelo neurônio vencedor, menor a especialização. Daí porque a topologia quadrada é a que oferece menor velocidade de especialização, pois, nela, o neurônio vencedor influencia uma maior quantidade de neurônios da Camada de Saída. Por sua vez, quando o Operador da Rede faz a opção pela topologia em cruz, decide então pela maior velocidade de especialização, pois nela a influência do neurônio vencedor em relação a seus vizinhos é mínima. A topologia em estrela oferece velocidade intermediária de especialização.

No caso anteriormente ilustrado pela Figura 10, quando restou vencedor o neurônio 34, afinal, quem são os seus vizinhos? Supondo que a topologia da

vizinhança foi definida como sendo em cruz e o raio da vizinhança como sendo de 2 unidades, então, os vizinhos são os dois neurônios imediatamente abaixo do vencedor, ou seja, os neurônios 14 e 24, os dois imediatamente acima, 44 e 54, os dois logo à esquerda, 32 e 33, e o único à direita, o 35, uma vez que não há um segundo neurônio imediatamente à direita do vencedor já que ele se encontra bem próximo à borda direita da Camada de Saída de dimensões 5x5.

Comparativamente, se a topologia da vizinhança tivesse sido definida como em estrela, além dos neurônios já mencionados seriam igualmente ajustados, em virtude da vizinhança, os neurônios 12, 23, 25, 43, 45 e 52. Convém observar, uma vez mais, que, em decorrência do neurônio vencedor estar muito próximo da borda direita da Camada de Saída, apenas um neurônio é favorecido na direção nordeste e um na direção sudeste, pois não existem dois neurônios nestas direções.

Em suma, a difusão do estímulo, para os neurônios próximos, deve ser parametrizada pelo Operador da Rede; tanto em relação às direções em que ocorrerá a difusão, quanto em relação ao raio, ou seja, à distância que a difusão alcançará.

3.9 Critérios de parada

Mas até quando os pesos serão ajustados? Não indefinidamente. Há casos em que notadamente não é mais necessário prosseguir com o aprendizado. Quando, por exemplo, não estiverem mais acontecendo alterações nos pesos a partir de novas entradas, não há mais sentido em prosseguir com o treinamento. O mesmo vale para quando as entradas se tornam constantes, situação em que a continuidade do treinamento novamente não traz novo aprendizado para a rede. Em contrapartida, e quando estas situações não ocorrerem? Até quando o treinamento prosseguirá?

Para evitar que o treinamento da Rede Neural seja realizado indefinidamente, o Operador da Rede, como conhecedor da área de negócio que deve idealmente ser, estabelece critérios, a partir dos quais, a rede será considerada suficientemente treinada e o treinamento será interrompido, ficando concluída a aprendizagem com a consequente submissão dos dados para efetiva classificação pela rede. É importante destacar que o adequado ajuste dos critérios de parada,

descritos a seguir, pode otimizar sobremaneira o resultado alcançado pelo treinamento. São eles:

- a) o **número de entradas** apresentadas à rede, ou seja, a quantidade de dados processados;
- b) o **número de ciclos** (ou iterações) processadas, ou seja, a quantidade de vezes em que o banco de dados foi integralmente processado;
- c) o **lapso de tempo** decorrido de treinamento;
- d) o valor da **taxa de aprendizagem alfa (α)**, que é decrementada com o tempo até atingir valores bem pequenos, quando o aprendizado torna-se insignificante;
- e) a **distância global entre os neurônios**, ou seja, a distância euclidiana entre os neurônios de entrada e da Camada de Saída que, quando atingem níveis muito baixos, já embutem alto nível de especialização, desfavorecendo a relação custo/benefício para um novo aprendizado; e
- f) a **convergência** da rede, ou seja, a repetição insistente de ganho como vencedor, sempre para o mesmo neurônio.

Este assunto é abordado de maneira mais aprofundada no tópico 4.7.3, relativo à Definição dos critérios de parada.

3.10 Os mapas auto-organizáveis de Kohonen

Teuvo Kohonen, pesquisador finlandês da Universidade de Helsinque, desenvolveu nos anos 80, uma rede neural baseada em aprendizado não-supervisionado conhecida como **Mapas Auto-organizáveis de Kohonen**, ou, simplesmente, Rede Neural de Kohonen e que é aqui objeto de nosso estudo e aplicação, uma vez que foi identificada, ao ser utilizada no funcionamento da rede, especial habilidade, para lidar com a elevada duplicidade de dados encontrada nos cadastros imobiliários, onde existe uma grande quantidade de imóveis com características semelhantes. Vide, por exemplo, um conjunto habitacional com 300 unidades de mesma área e com as mesmas características construtivas. A informação sobre tais imóveis difere, unicamente, pelo endereço, sendo redundante em relação a todos os outros dados.

A Rede Neural de Kohonen baseia-se no mapa topológico presente no córtex cerebral, onde neurônios topologicamente próximos tendem a responder de forma semelhante a determinados estímulos.

São componentes básicos da Rede Neural de Kohonen as camadas de Entrada e de Saída, esta última também conhecida como Camada de Competição, pois nela seus neurônios irão competir entre si, a fim de determinar qual deles possui maior proximidade da entrada apresentada à rede, significando maior afinidade entre suas características. Todo neurônio da Camada de Competição possui uma conexão virtual com cada neurônio da Camada de Entrada. A força dessa conexão é representada por pesos inicializados randomicamente com valores baixos entre 0 e 1, e que serão ajustados ao longo do treinamento.

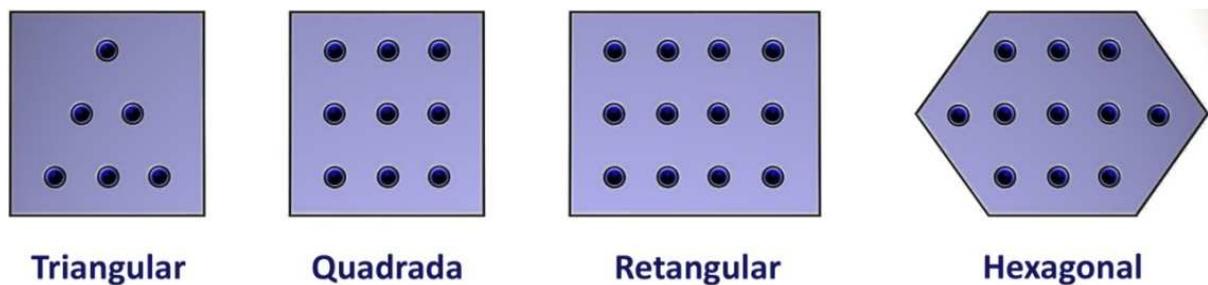


Figura 11 – Exemplos de topologias das Camadas de Saída

Considerando uma rede bi-dimensional, não lhe é imposta qualquer topologia específica, permitindo que os neurônios da Camada de Saída fiquem dispostos de maneira triangular, quadrada, retangular, hexagonal ou outra, como ilustrado na Figura 11.

Na fase de treinamento da Rede Neural de Kohonen, na medida em que ela é estimulada com novas entradas, passa a ocorrer uma competição entre os neurônios da Camada de Saída, a fim de identificar qual deles possui maior afinidade com a entrada. O vencedor terá seus pesos ajustados, tanto quanto seus vizinhos mais próximos. O treinamento prosseguirá até que a rede seja considerada treinada em virtude do atingimento de um ou mais critérios de parada definidos pelo Operador da Rede, como será visto no tópico 3.9.

Após o aprendizado, tem-se a fase de classificação, quando todas as entradas são apresentadas à rede e agrupadas de acordo com análise de similaridade baseada no aprendizado. São então formados os *clusters* que

correspondem à dispersão dos grupos, quando ficam próximas as entradas similares.

A seguir é apresentado o algoritmo básico que descreve o funcionamento da Rede Neural de Kohonen desde o pré-processamento até a classificação final dos dados de entrada. O correspondente fluxograma é mostrado na Figura 12:

```

1   --- Fase 1 : pré-processamento
2   Lê bases de dados;
3   Filtra valores inválidos;
4   --- Fase 2 : Treinamento
5   Parametriza a rede
6   Dimensiona a Camada de Saída
7   Enquanto existirem entradas e não tiver sido alcançado
   qualquer critério de parada
8     Recebe entrada para treinamento
9     Normaliza a entrada
10    Calcula distâncias euclidianas
11    Identifica o neurônio vencedor
12    Ajusta peso do neurônio vencedor
13    Ajusta peso dos vizinhos do neurônio vencedor
14  Fim do Enquanto
15  --- Fase 3 : Classificação
16  Enquanto existirem entradas normalizadas
17    Recebe entrada normalizada para classificação
18    Calcula distâncias euclidianas
19    Classifica a entrada, identificando o neurônio vencedor
20  Fim do Enquanto

```

Passa-se, então, a descrever como se desenvolve cada etapa, discorrendo detalhadamente sobre a implementação da rede.

A primeira fase, denominada de **pré-processamento**, tem por objetivo fazer uma análise prévia dos dados de entrada, a fim de evitar o retardamento do treinamento causado por dados inexpressivos, tanto quanto distorções no aprendizado causadas por dados incoerentes. Por exemplo, entradas repetidas em decorrência de cadastros duplicados devem ter suas duplicidades removidas, mantendo-se apenas a primeira entrada. De forma análoga, uma entrada em um cadastro imobiliário que indique que a área de um determinado terreno é nula, deve ser desprezada.

Dados nulos flagrantemente caracterizam incoerências cadastrais em muitos casos, tal como acontece com o exemplo das áreas de terrenos anteriormente mencionado, haja vista que não existem terrenos sem área. Assim, tais entradas deverão ser desprezadas já no pré-processamento, fazendo com que não interfiram negativamente em treinamentos futuros da Rede Neural.

A segunda fase, denominada **treinamento** ou **aprendizado**, é o momento mais delicado do funcionamento da rede. À primeira vista, pode parecer que os dados de entrada serão apresentados à rede e, como que, em um passe de mágica, serão corretamente identificados e agrupados sem qualquer equívoco para os padrões cerebrais humanos de pensamento. Ora, o raciocínio, com o qual se pretende dotar a rede, advém de um mecanismo que deve ocorrer a partir da percepção da relação entre os estímulos externos, as entradas, e o conhecimento já adquirido. Todavia, para acontecer a incorporação de novo conhecimento, a rede necessita de regras e orientações que a direcionem neste sentido, qual seja, o de maximizar a percepção e, conseqüentemente, o acúmulo do conhecimento.

A intervenção humana pode favorecer grandemente o processo de aprendizado através da **parametrização** da Rede Neural, o que deve anteceder o treinamento. A definição de tais parâmetros deve ser realizada por alguém a quem convencionou-se chamar de **Operador da Rede Neural**, Operador da Rede ou simplesmente Operador, a quem será confiada a meticulosa tarefa de definir, dentre outros parâmetros, os seguintes:

- a) os neurônios que irão compor a Camada de Entrada;
- b) a dimensão da Camada de Saída;
- c) os critérios de parada (já amplamente comentados no tópico 3.9);
- d) a forma como devem ser normalizadas as entradas; e
- e) a topologia e o raio da vizinhança.

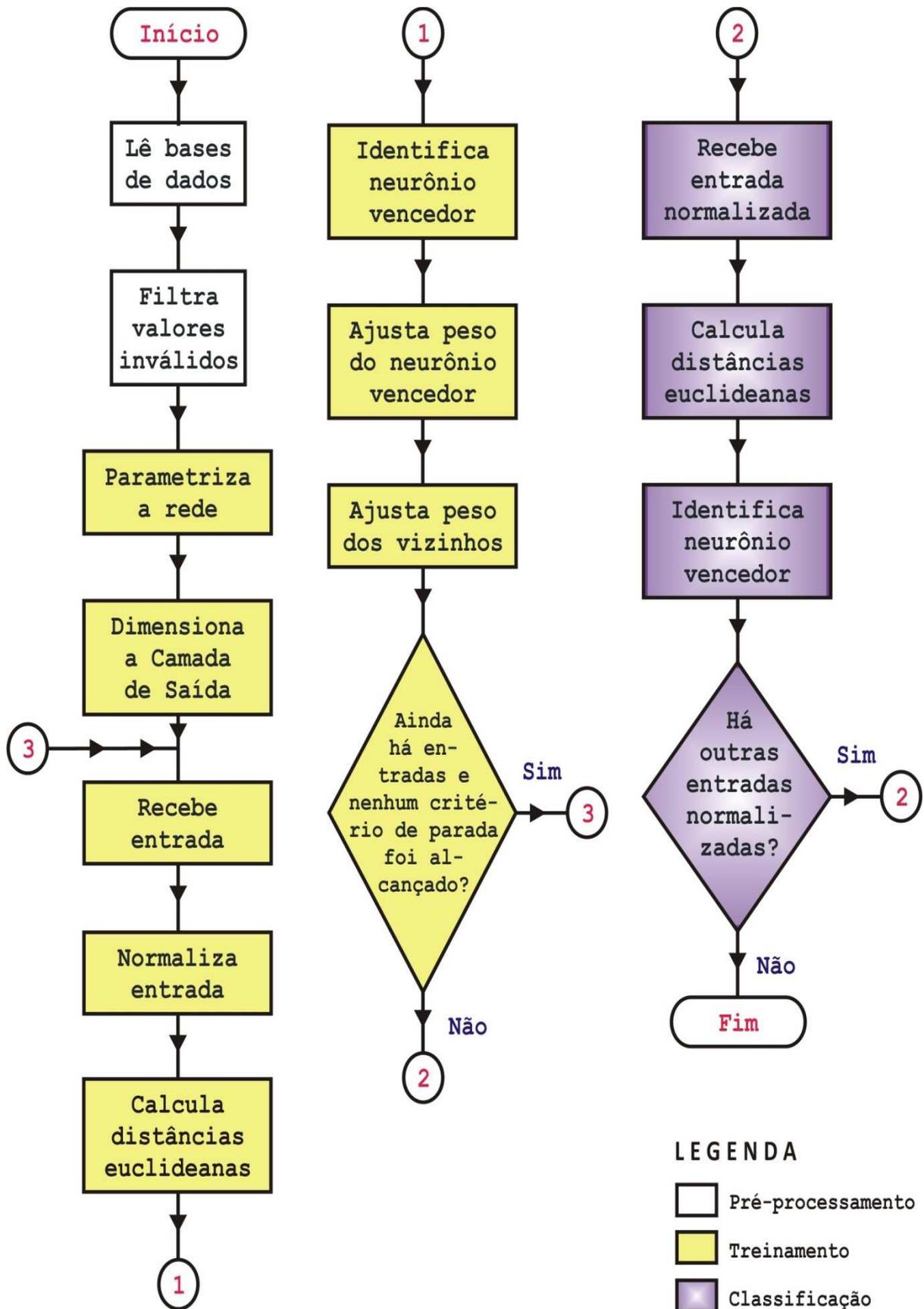


Figura 12 – Fluxograma do funcionamento da Rede Neural de Kohonen

Cada entrada lida, a partir do banco de dados já filtrado no pré-processamento, pode conter uma ou mais características dos dados. Por exemplo, considerando um cadastro imobiliário, uma entrada lida pode trazer informações sobre a área do terreno, a área edificada, o tipo do imóvel (territorial ou edificado), a classificação arquitetônica (casa, apartamento, galpão, loja...), o uso específico (residencial ou não residencial), o valor venal ou outros componentes dos imóveis. Caberá ao Operador da Rede Neural definir quais destas características formarão os neurônios que irão compor a Camada de Entrada. Desta forma, se julgar conveniente, o Operador poderá definir que serão classificadas todas as entradas, considerando apenas os neurônios *área do terreno*, *área edificada* e *classificação arquitetônica* (vide exemplo na Tabela 3 na página 59).

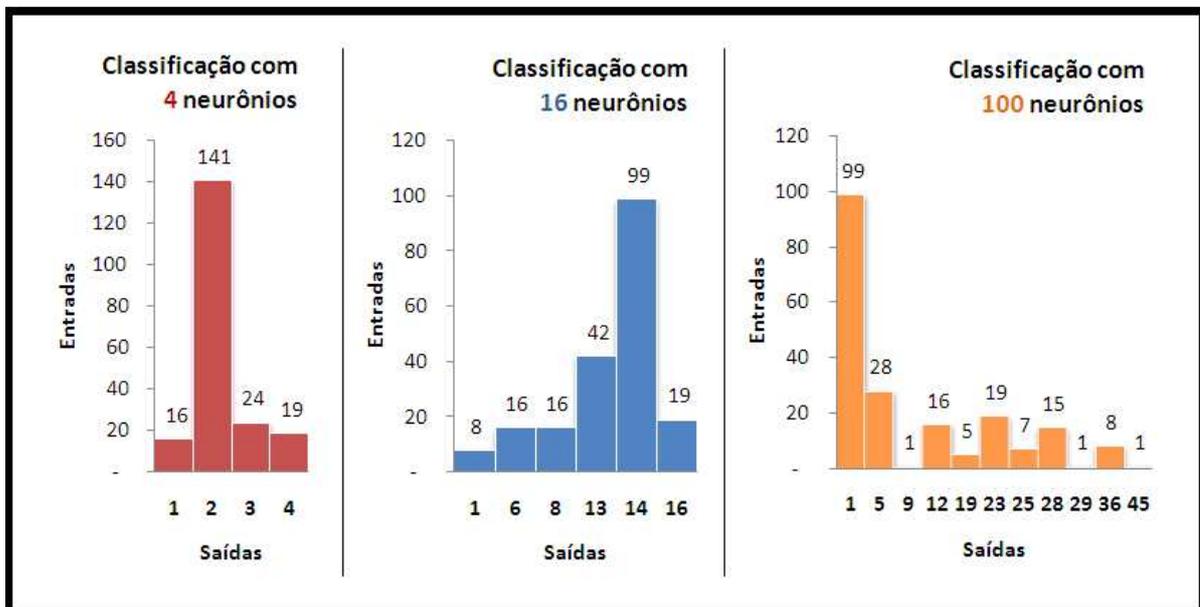


Figura 13 – Especialização da RNA com base na quantidade de neurônios da Camada de Saída

A Camada de Saída pode conter de dois a muitos neurônios. Esta quantidade definirá a dimensão da Camada de Saída. Cada um destes neurônios poderá formar um agrupamento específico que atrairá um determinado conjunto de entradas que apresentem características semelhantes. Se a Rede Neural possuir uma Camada de Saída com topologia quadrada e dimensão 2x2, apenas poderá reconhecer e classificar os dados de entrada em quatro grupos distintos, gerando considerável generalização. Em contrapartida, na medida em que a quantidade de neurônios da Camada de Saída aumenta, melhora também a sua especialização.

A Figura 13 mostra um quadro comparativo que permite inferir a maior especialização da RNA com maior quantidade de neurônios na Camada de Saída, em detrimento daquela com menor quantidade. A rede gerada com 16 neurônios na Camada de Saída, estruturada em uma matriz 4x4 em topologia quadrada, gerou 6 grupos distintos, ou seja, 6 neurônios desta camada tiveram a eles associadas uma ou mais entradas. Já a rede quando definida com 100 neurônios na Camada de Saída, gerou 11 grupos para a mesma amostra de 200 dados de entrada, provenientes de um determinado banco de dados imobiliário. Os dados mencionados foram organizados na forma disposta na **Erro! Auto-referência de indicador não válida..**

Tabela 4 – Grau de especialização em função da dimensão da Camada de Saída

Situação	Neurônios na Camada de Entrada (entradas)	Neurônios na Camada de Saída (saídas)	Grupos formados	Grau de Especialização
1	200	4	4	Baixo
2	200	16	6	Média
3	200	100	11	Alto

Uma vez definidos os neurônios que irão compor a Camada de Entrada, bem como a dimensão da Camada de Saída, a RNA será submetida ao aprendizado, através do qual, acumulará conhecimento, a fim de poder, na fase seguinte, classificar as entradas em agrupamentos semelhantes. No entanto, devem ser considerados alguns fatores que configurem o treinamento, para que aconteça em tempo aceitável e com o nível de especialização desejado. Como a Rede Neural Artificial é capaz de lidar com estruturas extremamente complexas, com dezenas ou mesmo centenas de neurônios na Camada de Entrada e quantidade semelhante de neurônios na Camada de Saída, o tempo necessário, para que a rede chegue ao ponto de não mais acumular conhecimento em virtude da grande especialização que alcançou, pode ser demasiadamente grande.

Para limitar o tempo de treinamento a uma duração razoável, o Operador deve estipular parâmetros, para que a RNA, ao alcançá-los, entenda como aceitável o nível de aprendizado ao qual chegou e interrompa o treinamento, passando imediatamente à classificação das entradas, a fim de separá-las em grupos semelhantes. Estes parâmetros são conhecidos como critérios de parada, pois, no momento em que um ou mais deles for alcançado, a rede é considerada treinada e para o treinamento. Devem ser meticulosamente ajustados de acordo com o

conhecimento do negócio dominado pelo Operador da Rede, por vezes sendo necessário defini-los a partir dos resultados obtidos em treinamento anterior.

Os critérios de parada podem ser muitos, de acordo com a necessidade e as particularidades de cada caso, mas os principais e que também foram adotados neste trabalho são: a quantidade de entradas, o número de iterações processadas, a duração do treinamento, o valor alcançado pela taxa de aprendizagem, a distância máxima entre os neurônios de entrada e de saída e a convergência da rede. Como o assunto já foi amplamente comentado no tópico 3.9, recomenda-se que seja consultado para maiores referências.

A parametrização do funcionamento da RNA prossegue com a definição da forma como devem ser normalizadas as entradas. No tópico 3.5 também já foram explicitadas exaustivamente as maneiras de tratamento dos dados para fins de normalização, tanto, quando os dados têm natureza de variáveis discretas, quanto em situações em que têm natureza de variáveis contínuas. O Operador da Rede, uma vez mais, como conhecedor do negócio que deve ser, deve regular a parametrização para um melhor rendimento e performance da RNA. Novamente recomenda-se a consulta ao tópico mencionado, para maiores esclarecimentos.

Por fim, o processo de planejamento do funcionamento da rede a ser executado pelo Operador da Rede encontra a necessidade de definir a topologia e o raio da vizinhança em relação ao neurônio vencedor, para dirigir o ajuste de pesos da maneira mais apropriada e que favoreça ao treinamento e à classificação eficaz das entradas apresentadas à rede. Anteriormente, no tópico 3.8, foram explicados tanto a topologia quanto o raio aplicáveis à vizinhança de uma RNA, pelo que convém sugerir uma revisita ao tópico mencionado, caso se deseje maiores detalhes sobre o assunto.

Para resumir, então, a fase de treinamento, tem-se ao seu final um conjunto de pesos associados a cada neurônio da Camada de Saída. Tais pesos foram evoluindo na medida em que as entradas foram sendo apresentadas à RNA, uma a uma, e novo aprendizado foi sendo agregado à rede em virtude da análise de semelhança entre o conjunto de neurônios da Camada de Entrada apresentada e o conjunto de neurônios da Camada de Saída definida.

A terceira e última fase, denominada **classificação**, é a consumação da efetiva aplicação da Rede Neural Artificial, equivalendo ao seu objetivo final. Na classificação, todas as entradas são apresentadas à rede tal como na fase de

treinamento. É identificado o neurônio vencedor e a entrada é imediatamente classificada na posição do neurônio vencedor, ou seja, no grupo ao qual pertence o neurônio vencedor. Ao contrário da fase de treinamento, na fase de classificação não mais existe ajuste de pesos dos neurônios da Camada de Saída. Desta forma, uma a uma, as entradas vão sendo classificadas. Ao final, os dados estão inteiramente classificados e qualquer nova entrada pode ser apresentada para imediata classificação, sem a necessidade de novo treinamento da Rede Neural.

Para efeito de melhor visualização, os dados podem ser dispostos de forma espacial e apresentados graficamente. Podem também ser listados de acordo com o grupo ou neurônio de Saída ao qual pertencem, para fins de maiores estudos ou até mesmo para nova submissão à rede e um novo treinamento específico apenas para os dados pertencentes ao grupo desejado. Seria a especialização da especialização, ou a classificação da classificação.

3.11 Considerações finais

A pesquisa e a tecnologia dela decorrente que foram desenvolvidas desde a década de 40 por McCulloch e Pitts, quando criaram as Redes Neurais Artificiais, vêm sendo aprimoradas por inúmeros cientistas e pesquisadores até a atualidade, passando pela criação dos Mapas Auto-organizáveis de Kohonen, e trazendo relevantes avanços para a computação, quando viabilizaram uma forma diferente e eficaz de agregar conhecimento fazendo uso de computadores. Com a capacidade de aprender, a partir de dados externos, com ou sem a interferência humana, deu-se novo alento à habilidade de classificar objetos e até mesmo comportamentos, a fim de identificar agrupamentos semelhantes tal como se demonstrou até aqui.

Conclui-se, então, que o mecanismo oferecido pelas RNA para o aprendizado e a posterior classificação de dados torna-se capaz a partir da aplicação de conceitos matemáticos, procedimentos e implementação de médio nível de complexidade, apresentando ótima relação custo/benefício.

Com a alocação de um Operador da Rede com mediano nível de entendimento em informática, a disponibilização de um computador com performance compatível com o uso doméstico, e a utilização de *software* adequado (em nosso caso o Analisador do Solo Urbano, desenvolvido pelo mestrando

especialmente para a demonstração do uso prático das Redes Neurais Artificiais), é possível classificar dados cujas semelhanças são buscadas sem a necessidade de qualquer conhecimento prévio da relação entre eles. É conveniente, apenas, que o Operador seja conhecedor do negócio, ou seja, dos dados a serem classificados, a fim de otimizar o treinamento por meio da adequada parametrização da RNA.

O resultado da classificação, conjuntamente com os métodos tradicionais de seleção de dados, facilita a identificação de distorções, diminuindo o erro na adoção de ações tendentes a corrigir o cadastro, seja ele imobiliário ou outro. O aumento da eficiência, da eficácia e da efetividade na tomada de decisões, por si só, já equivale a sensível benefício para o trabalho.

4 ESTUDO DE CASO

4.1 Considerações iniciais

O município de Fortaleza foi escolhido por ter uma grande variedade imobiliária, estar desatualizado quanto ao cadastro existente e por ter dado autorização para o uso dos dados. Há pelo menos 6 anos, não é feita revisão em massa nas características dos imóveis, nem publicação de nova Planta Genérica de Valores Imobiliários, ficando desatualizados os valores das parcelas de solo, como também das edificações cujas características com impacto tributário foram modificadas sem a concomitante informação ao Fisco municipal.

Dos mais de 500 mil imóveis tributados existentes no Cadastro Imobiliário municipal, no presente estudo de caso trabalhou-se com uma amostra de 10994 imóveis que compõem o distrito 15, relativo ao bairro da Aldeota. Nele foram aplicadas as técnicas de clusterização desenvolvidas neste trabalho. Para que fosse possível a utilização dos dados, a Secretaria de Finanças do Município (SEFIN) autorizou sua utilização, como gestora que é dos dados em questão, no sentido de que a técnica restasse demonstrada com dados reais e atuais, resguardados aqueles protegidos por lei.

As especificações de um Cadastro Imobiliário Municipal usualmente se encontram dispersas em vários bancos de dados. No caso do município de Fortaleza, tal cadastro é gerido e mantido pela Secretaria de Finanças do Município (SEFIN), através da Célula de Gestão do IPTU (CGIPTU) com o suporte da Célula de Tecnologia da Informação (CELTU). O principal arquivo de controle imobiliário, o cadastro de terrenos e edificações, permite registrar os imóveis através da identificação de suas características básicas.

Pela notada escassez de recursos humanos, técnicos e financeiros, a atualização continuada dos dados referentes a cada um dos mais de 600 mil imóveis cadastrados pelo Fisco é deveras dificultada, tornando-se pouco eficaz e de abrangência limitada, o que inviabiliza a correta tributação das unidades imobiliárias, até porque o próprio contribuinte, via de regra, não cumpre a obrigação legal de informar ao Fisco, quando existem alterações realizadas nas características do

imóvel, tais como a construção de um segundo pavimento ou a alteração de características construtivas em geral.

A defasagem entre o valor tomado como base de cálculo para a tributação do IPTU a cada início de exercício fiscal, mais precisamente em todo 1º de janeiro de cada ano, e o valor real dos imóveis para o mercado imobiliário, toma dimensões assustadoras, fazendo com que o imposto cobrado do contribuinte seja um dos mais baixos dentre as capitais da Região Nordeste do país.

Casos de inacreditável subavaliação têm origem, ainda, na falta de atualização dos valores dos terrenos conforme suas localizações, o que deveria ser feito periodicamente através da Planta Genérica de Valores Imobiliários (PGVI), a ser aprovada pelo Poder Legislativo Municipal a partir de estudos técnicos e proposta encaminhada pelo Poder Executivo. Notadamente, por se tratar de imposto com grande impacto social e político, são encontrados inúmeros empecilhos sempre que sua atualização é tratada tanto no âmbito do Executivo, como no Legislativo, uma vez que a abrangência de seus ajustes alcança praticamente toda a população, mesmo quando se reconhece, de forma incontestável, a sua subavaliação.

Muitas vezes, a não aceitação de qualquer proposta de ajuste do imposto tem estreita ligação com a crença dos contribuintes de que o tributo cobrado não trará, ato contínuo, benefícios para a sociedade. Tamanha desatualização faz com que alguns imóveis sejam avaliados, para efeito de cobrança do IPTU, por menos de 20% do seu real valor no mercado imobiliário, significando que, nestes casos específicos, a tributação deveria estar ocorrendo sobre uma base de cálculo até cinco vezes maior, gerando, certamente, sensível aumento na arrecadação municipal. E, quanto mais se posterga a atualização dos valores, mais desatualizados eles se tornam em virtude da valorização imobiliária natural, especialmente em cidades com grande especulação imobiliária, como é o caso de Fortaleza.

4.2 Equipamento utilizado

O preparo dos dados, o treinamento da Rede Neural e a classificação foram todos modelados e executados em um microcomputador comum, possuindo as especificações contidas do Quadro 3:

Item	Dispositivo	Especificação
1	Placa mãe	VS Company G31T-M
2	Processador	Intel Celeron 1.8 GHz
3	Memória RAM	2 GB DDR2 800 MHz
4	Controladora de disco	Intel 82801G Serial ATA
5	HD	Samsung SATA 500 GB (ST3500418AS)
6	Monitor	LCD Samsung 20" (mod. 2063UW) máx. 1680x1050
7	Adaptador de vídeo	Intel G31/G33 Express Chipset Family

Quadro 3 – Especificação do equipamento utilizado no estudo de caso

Ainda que o processador Intel Celeron 1.8 GHz utilizado não represente o estado-da-arte, o computador mostrou-se suficientemente rápido para os testes realizados, sendo capaz de treinar a Rede Neural da amostra equivalente ao bairro da Aldeota, com aproximadamente 11 mil imóveis, em menos de um minuto, levando igual tempo para classificar os dados. Maiores detalhes sobre os experimentos do treinamento e da classificação estão expostos no tópico 4.7.5.

4.3 A ferramenta de apoio à decisão

Em virtude da enorme quantidade de dados a serem processados e tratados, foi necessário criar um *software* específico com o objetivo de otimizar a obtenção de informações a partir dos bancos de dados do Cadastro Imobiliário, informações essas que permitissem a identificação de distorções tributárias causadoras de perda de recursos para os cofres municipais.

O *software* apresentado na Figura 14 e batizado de Analisador do Solo Urbano, ou simplesmente Analisador, foi integralmente desenvolvido pelo mestrando, utilizando uma linguagem de programação acessível, *Visual Basic*, mesclando conhecimentos de programação com aqueles relativos à gestão tributária do solo urbano no município de Fortaleza, onde o mestrando ocupa o cargo de Auditor do Tesouro Municipal, exercendo há mais de quatro anos a gerência da Célula de Gestão do IPTU da Secretaria de Finanças do Município.

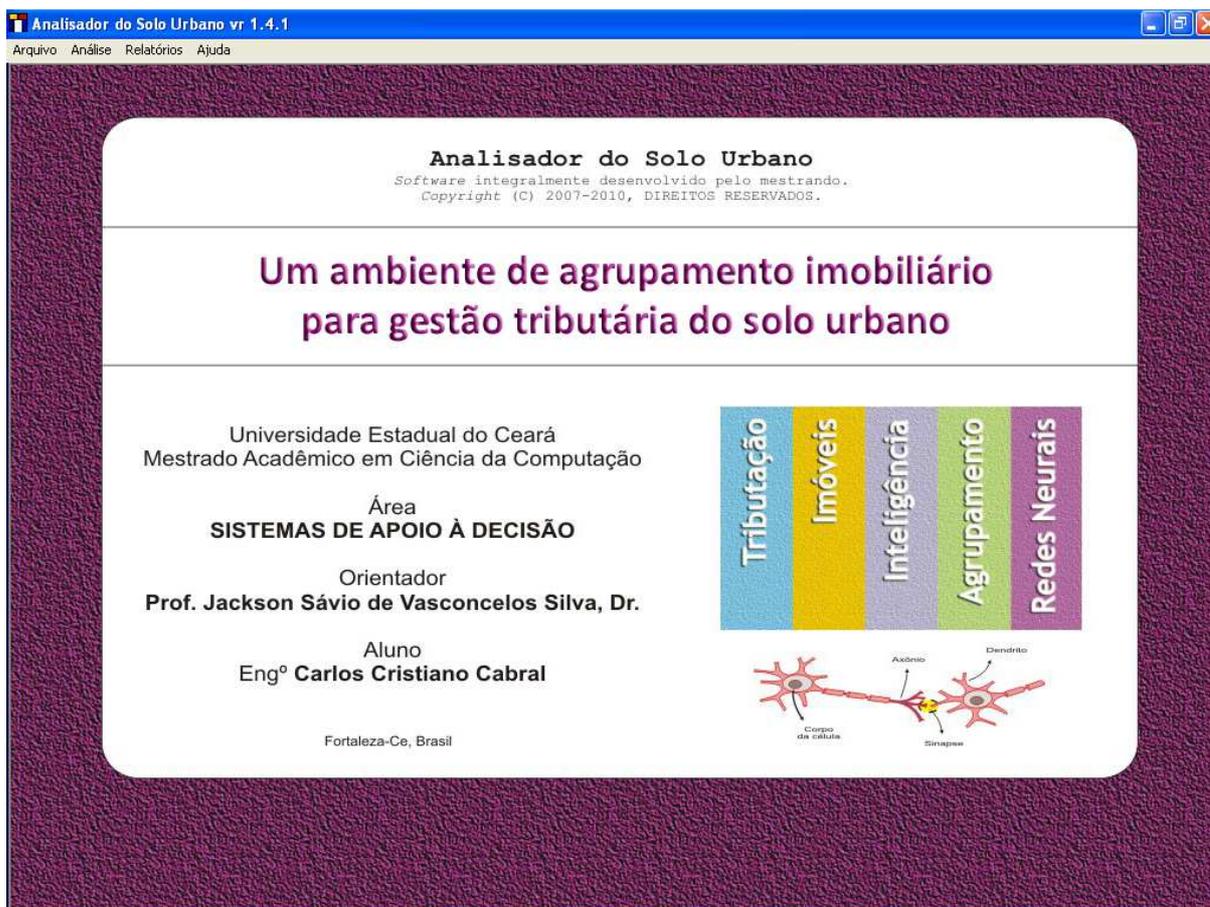


Figura 14 – Tela inicial do *software* Analisador do Solo Urbano em sua versão 1.4.1.

As mais de 10 mil linhas de código do programa fonte e de documentação escritas ao longo do Mestrado Acadêmico em Ciência da Computação, da Universidade Estadual do Ceará (UECE), deram vida a um *software* desenvolvido com visão eminentemente acadêmica, tanto que seus arquivos de entrada, bem como seus relatórios, todos são legíveis por editores de texto comuns. Inclusive a metodologia de aprendizado é explicada passo-a-passo em um relatório detalhado no qual são descritos, além dos parâmetros de configuração da Rede Neural, os cálculos das distâncias euclidianas, a forma de definição do neurônio vencedor e, por fim, o mecanismo de ajuste de pesos do vencedor e de seus vizinhos. Todo o código, bem como a documentação foram escritos pelo mestrando, quem foi responsável, ainda, pela análise do sistema e pelas ilustrações e desenhos, tanto os constantes no software final, como neste documento.

Após a implementação e, no estado em que se encontra, o *software* além de demonstrar larga aplicação na tributação imobiliária, fez ver que com as devidas adaptações, poderia ser facilmente aplicável à classificação de dados dos mais

diversos, uma vez que a metodologia de agrupamento é válida para quaisquer variáveis, dos mais diferentes tipos, discretas ou contínuas, já que são normalizadas, o que as torna comparáveis entre si. Assim, já se vislumbra o desenvolvimento de uma versão com variáveis plenamente configuráveis pelo usuário, e que seria de grande utilidade para o agrupamento de dados semelhantes tanto na esfera pública, quanto na esfera privada.

O Analisador desenvolve e implementa as rotinas utilizadas ao longo deste trabalho, tais como:

- a) Importação e conversão dos dados brutos;
- b) Segmentação dos dados em distritos;
- c) Análise estatística;
- d) Geração de relatórios gerais e específicos por distrito e por característica dos imóveis;
- e) Pesquisa de dados com visualização espacial dos resultados;
- f) Listagem de quantitativos de imóveis por quadra, para o filtro selecionado;
- g) Navegação pelas quadras plotadas, com identificação dos quantitativos apurados; e
- h) Classificação de imóveis através de Redes Neurais Artificiais.

O banco de dados original em formato DATAFLEX foi importado e convertido para o formato CSV (*Comma Separated Values*), ou seja, texto separado por vírgulas. Com a rotina de Análise do Cadastro do IPTU, o Analisador segmenta os dados dos 77 distritos em arquivos separados, gerando arquivos de largura fixa, a fim de que o processamento seja otimizado, uma vez que o objetivo do trabalho é classificar dados de distritos específicos através das RNA. Ainda assim, o *software* também implementa rotina de pesquisa e visualização espacial de todo o município de uma só vez.

Como o Analisador foi inicialmente concebido para fins didáticos, ainda que posteriormente tenha sido exclusivamente por nós utilizado no próprio planejamento tributário, para fins de direcionamento na execução de trabalhos específicos na Célula de Gestão do IPTU (CGIPTU), não se adotou qualquer padrão de dados mais eficiente e seguro, cuidando-se ainda de fornecer, tanto quanto possível, elementos de estudo da tecnologia de Redes Neurais implementada, como

no caso do treinamento, quando todo o processo é explicado tecnicamente, passo-a-passo, em arquivo texto descritivo das operações realizadas.

A velocidade no acesso aos dados findou reduzida, devendo oportuna e futuramente ser melhorada com o desenvolvimento de uma versão profissional para o *software* quando, por exemplo, pode ser adotada o padrão *Structured Query Language* (SQL), para acelerar a pesquisa, conferindo-lhe ainda maior segurança. Vale ressaltar que, ainda que a rede utilize o formato texto (TXT), os tempos de treinamento e classificação obtidos, para um único distrito pela Rede Neural, atingem níveis plenamente aceitáveis, mesmo em equipamentos simples, como o descrito no tópico 4.2.

Paulatinamente, o que era para ser apenas um programa que demonstrasse a classificação de dados, através de uma Rede Neural, evoluiu para um ambiente complexo de pesquisa, visualização, análise e classificação de dados com aplicação de metodologia científica para a identificação de distorções tributárias, com utilidade prática já comprovada no ambiente profissional no qual foi inserido. Infelizmente, em virtude da fragilidade dos dados em formato TXT, a utilização do *software* não pode ser ampla e livremente difundida, antes que sejam adotados os necessários procedimentos de proteção dos dados contra acessos não autorizados, como asseveram Priddy e Keller (2005, p.23).

O primeiro passo ocorre, quando da análise dos dados. Nela, as características dos imóveis pertencentes a cada distrito são quantificadas em separado, quando, ao seu final, são criados relatórios de diversos tipos, tais como de totalização, de inconsistências e de imóveis com endereços incompletos.

4.4 Criação dos ambientes gráficos

A disposição do resultado do processamento de forma gráfica torna mais fácil e claro o entendimento. Desta forma, optou-se por adotar esta modalidade de visualização de informações, em detrimento de relatórios sob a forma de texto, meramente alfanuméricos. Foram desenhados no Analisador, para demonstrar o funcionamento das RNA, três ambientes gráficos que traduzem e ilustram os seguintes momentos:

- a) Mapeamento da pesquisa com o filtro selecionado;
- b) Competição dos neurônios da Camada de Saída;
- c) Classificação das entradas.

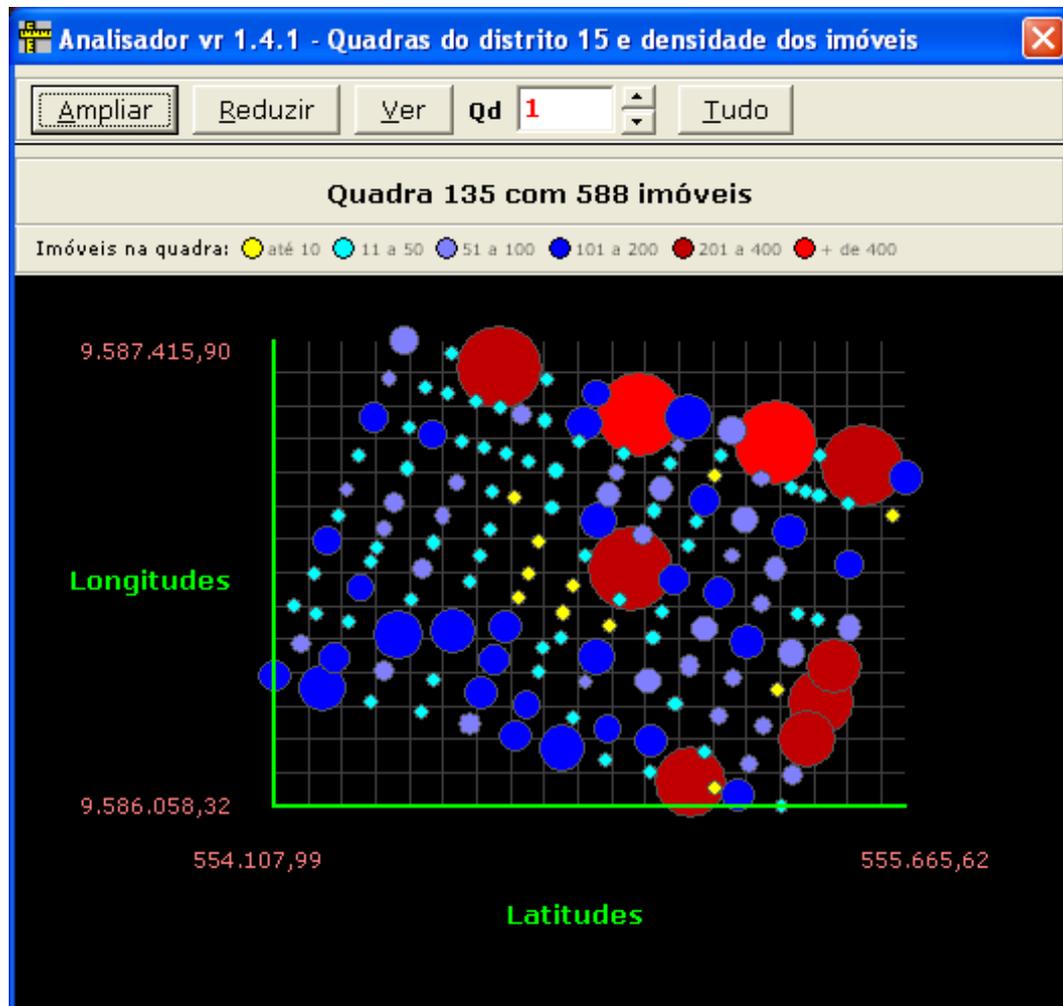


Figura 15 – Concentração de imóveis em geral no distrito 15 (Aldeota)

4.4.1 Ambiente gráfico de mapeamento da pesquisa

O ambiente de mapeamento da pesquisa é utilizado, para se localizar espacialmente e quantificar imóveis com uma determinada característica ou para mostrar a concentração deles, independente de seus tipos, em um determinado distrito ou em todo o município. No mapa, a concentração dos imóveis é graficamente representada por círculos de tamanhos diferentes cujas cores variam de acordo com 6 faixas de valores específicas, tal como ilustrado na Figura 16.

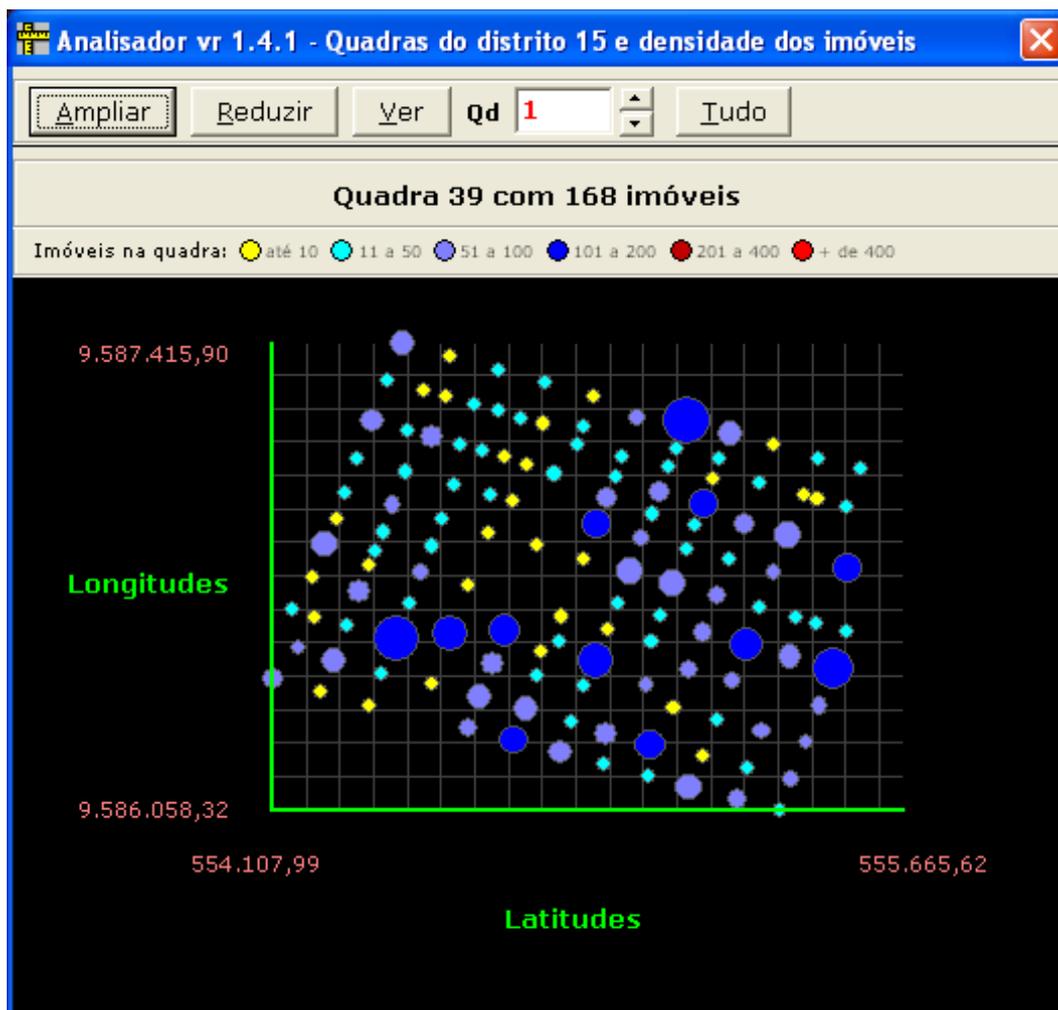


Figura 16 – Concentração de imóveis comerciais no distrito 15 (Aldeota)

A pesquisa da concentração de imóveis, que pode ser feita por distrito ou em todo o município, mostra-se de grande valia para o trabalho cotidiano da Administração Fiscal, uma vez que permite observar a distribuição da densidade de ocupação em função de cada tipo de imóvel, como, por exemplo, os terrenos. Mais ainda, facilita sobremaneira a localização de imóveis que satisfaçam a uma determinada condição, ficando fácil visualizar a concentração, digamos, de imóveis com áreas edificadas superiores a 1000 m².

Além da visualização espacial, a ferramenta desenvolvida lista os imóveis encontrados, em um arquivo no formato texto (TXT), separado por vírgulas, para facilitar a leitura a partir de planilhas eletrônicas, onde os dados poderão ser organizados inclusive a partir de outros campos listados. A listagem contempla campos como o número de inscrição do imóvel no cadastro imobiliário, o distrito, a quadra, a área do terreno (em *m*), a área edificada (em *m*²), a testada principal (em *m*), o tipo (territorial ou predial), o uso específico (residencial ou não residencial), a

classificação arquitetônica (barraco, casa, apartamento, galpão aberto, galpão fechado, sala, loja e outros), o valor venal cadastrado (em R\$), o valor do IPTU no exercício fiscal (em R\$) e outros.

Para que fosse estruturado o ambiente no qual os resultados das pesquisas seriam visualizados, foi necessária a montagem de um sistema de coordenadas bi-dimensionais onde se fez constar as latitudes no eixo das abscissas e as longitudes no eixo das ordenadas, com o fito de permitir utilizar as coordenadas geográficas do centro da quadra onde está localizado o imóvel, a fim de plotá-lo adequadamente no ambiente desenhado.

Para que se tornasse viável a estruturação do plano cartesiano com a posterior visualização espacial dos dados, tomaram-se as coordenadas geográficas da quadra mais à esquerda no distrito analisado, da quadra mais à direita, daquela localizada na parte mais extrema no limite superior, e, por fim, a coordenada da quadra localizada mais abaixo no limite inferior do município. Formado o quadrilátero encontrado pelos limites descritos, o plano cartesiano foi montado, a fim de que nele pudessem ser alocados os imóveis selecionados. Registraram-se, também, as coordenadas dos centros das quadras (centroides) do distrito sob análise. Desta forma, a visualização, por exemplo, de todos os imóveis residenciais tornou-se possível, fornecendo mais uma ferramenta de análise.

A opção de visualizar todas as quadras do distrito e nelas o indicativo da quantidade de imóveis contados com a característica selecionada é padrão. No entanto, é permitido ao usuário selecionar uma única quadra para análise ou ainda ampliar o mapeamento gerado para facilitar a visualização.

Como serviço adicional para o usuário do ambiente, cuidou-se de dar-lhe a faculdade de navegar sobre as quadras do distrito em análise, quando então é mostrado o quantitativo de imóveis com a característica selecionada para a quadra apontada. Ao posicionar o cursor sobre uma área vazia, é pesquisado o número da quadra mais próxima e a sua distância do ponto atual, o que é prontamente informado ao usuário.

O algoritmo que se segue descreve os passos utilizados para a montagem do ambiente gráfico de mapeamento da pesquisa, feita sobre um determinado distrito:

```

1   --- Fase 1 : Ajusta a origem do gráfico
2   Define a localização gráfica do X e do Y da origem,
    coordenada (0,0)
3   Define a extensão dos eixos X e Y
4   --- Fase 2 : Define eixos de coordenadas
5   Monta eixo das abscissas
6   Monta eixo das ordenadas
7   Mostra no eixo X a menor e a maior latitude
8   Mostra no eixo Y a menor e a maior longitude
9   --- Fase 3 : Monta grade de plotagem
10  Monta linhas horizontais
11  Monta linhas verticais
12  --- Fase 4 : Identifica centroides das quadras
13   $EscalaX = \frac{\text{Extensão do eixo X}}{(\text{Maior latitude} - \text{Menor latitude})}$ 
14   $EscalaY = \frac{\text{Extensão do eixo Y}}{(\text{Maior longitude} - \text{Menor longitude})}$ 
15  Para Q = 1 até Total de Quadras do Distrito
16       $X = X \text{ da origem} + (\text{LatitudeDaQuadra}(Q) - \text{MenorLatitude}) \times EscalaX$ 
17       $Y = Y \text{ da origem} - (\text{LongitudeDaQuadra}(Q) - \text{MenorLongitude}) \times EscalaY$ 
18      Plota centroide na coordenada (X,Y)
19  Próxima quadra Q

```

4.4.2 Ambiente gráfico de competição

O ambiente de competição é utilizado, para informar o neurônio vencedor e seus vizinhos durante o processo de treinamento. Tanto o vencedor, como os vizinhos afetados com o ajuste de peso para a topologia de vizinhança escolhida, são mostrados numa matriz quadrada com as dimensões escolhidas pelo Operador da Rede, para representar a Camada de Saída, na medida em que cada entrada é apresentada.

Na Figura 17, são mostradas duas situações em treinamentos distintos. No treinamento à esquerda, é mostrada a Camada de saída (ou de competição) definida na dimensão 10x10 para uma rede com vizinhança possuindo topologia em estrela e raio de 2 unidades. O neurônio vencedor pode ser identificado na cor mais clara (branca) e os vizinhos na cor mais escura (preta). Já no exemplo do outro treinamento, à direita, a Camada de saída (ou de competição) tem dimensão 6x6, a vizinhança da rede foi definida no formato em cruz, tendo um raio da vizinhança de 3

unidades. Uma vez mais o neurônio vencedor pode ser identificado na cor mais clara (branca) e os vizinhos na cor mais escura (preta).

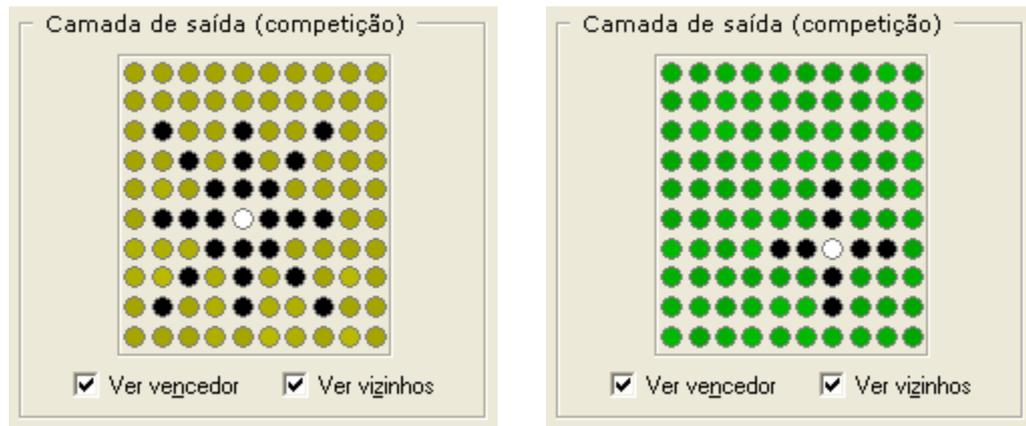


Figura 17 – Exemplos de mapeamento do neurônio vencedor e de seus vizinhos

4.4.3 Ambiente gráfico de classificação

O agrupamento de indivíduos semelhantes ganha corpo especial para fins de entendimento e percepção humanos, quando demonstrados de forma espacial. Neste sentido, mais que oferecer uma mera listagem dos neurônios agrupados, decidiu-se por criar um ambiente gráfico bidimensional que permitisse a visualização do estado final de classificação dos neurônios de entrada.

O ambiente de classificação é utilizado, para mostrar os agrupamentos formados pelas entradas que foram classificadas após a rede ser submetida ao treinamento. Demonstra o resultado final do processamento com o mapeamento das entradas utilizando os Mapas auto-organizáveis de Kohonen. Na Figura 18, os neurônios da Camada de Saída foram definidos em matriz 10x10 pelo Operador da Rede, aparecendo delimitados por 100 circunferências regulares. Destes, apenas alguns deram origem a agrupamentos. Os demais neurônios não demonstraram afinidade suficiente com as entradas e, portanto, não deram origem a nenhum grupo.

Em cada grupo pode ser observado visualmente dentro do círculo que o delimita, o nível de dispersão entre os indivíduos nele classificados. Uma maior dispersão indica que os dados selecionados no grupo em questão, possuem menor similaridade, pois estão mais distantes uns dos outros.

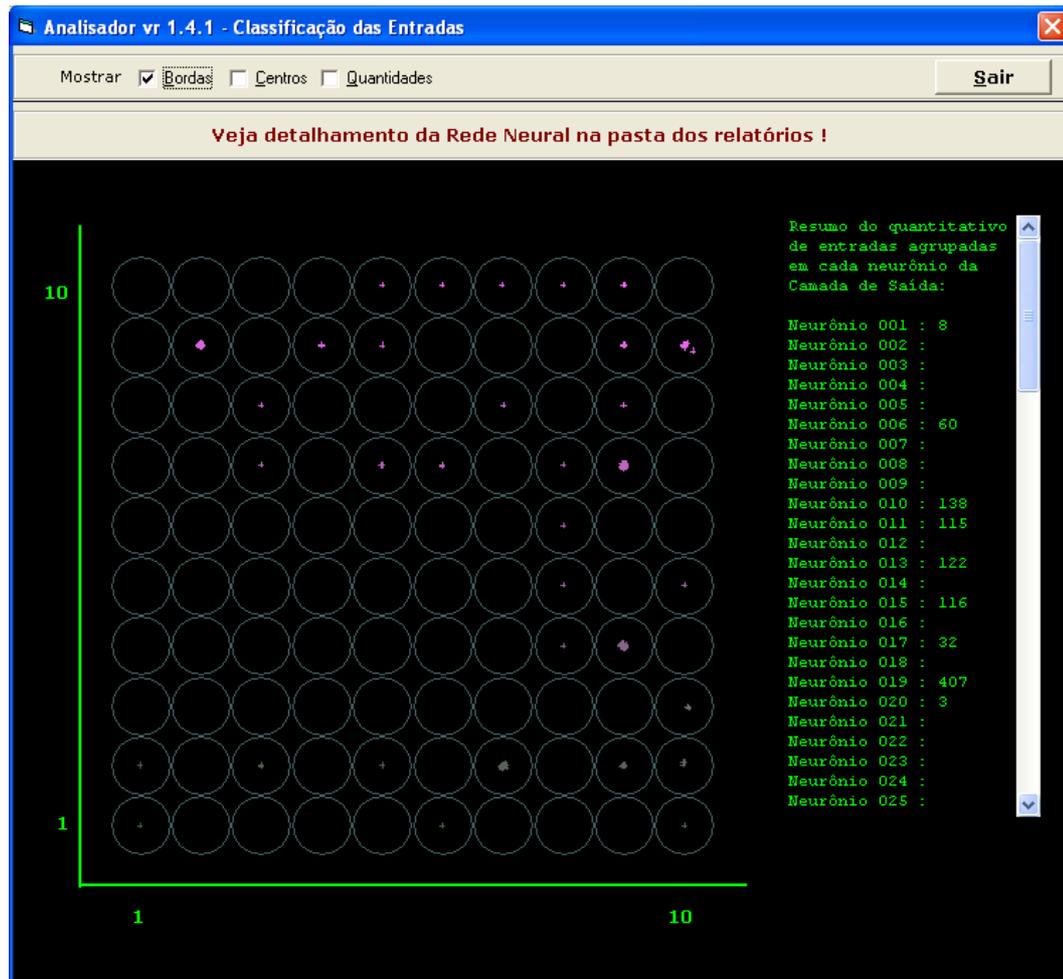


Figura 18 – Agrupamentos formados após a classificação das entradas

4.5 Obtenção de dados brutos

Abstraindo-se da existência de outros bancos de dados, tais como aqueles relativos ao cadastro de logradouros (ruas, avenidas, travessas...) e ao cadastro de contribuintes, o centro de nossa atenção volta-se ao cadastro dos imóveis, ou seja, ao Cadastro Imobiliário. Convém lembrar, também, que, com igual importância, relacionam-se ao Cadastro Imobiliário outros bancos de dados relevantes, como, por exemplo, o que contém os valores do metro quadrado do terreno para cada face de quadra no município, ou seja, sua Planta Genérica de Valores Imobiliários (PGVI).

As fontes de dados originárias, aí incluídos o cadastro principal com os dados dos imóveis, bem como outros arquivos auxiliares com informações diversas utilizados para o estudo de casos ora apresentado, foram aquelas constantes do

Quadro 4, a seguir. Vale frisar que diversos arquivos auxiliares e de controle foram criados ao longo do processamento da RNA, a fim de lhe dar o necessário suporte.

Banco de dados	Conteúdo
Cadastro de imóveis	Relação de imóveis e suas características
Tabela de campos	Descrição e tipos dos campos do cadastro de imóveis
Valores de campos	Descrição dos valores e pesos tributários válidos para os campos definidos na tabela de campos
Centroides das quadras	Localização geográfica dos centros de cada quadra

Quadro 4 – Fontes de dados utilizados no estudo de casos

Merece consideração o fato de que o arquivo contendo os centroides das quadras apresenta algumas coordenadas de quadra omissas, quando, por motivos diversos, a quadra em questão não existe na base de dados de centroides das quadras do distrito. Esta omissão não causa interferência na aplicação da Rede Neural Artificial implementada, uma vez que as coordenadas geográficas não foram utilizadas, quando da plotagem espacial dos agrupamentos.

No entanto, para a visualização espacial de uma determinada seleção, quando, por exemplo, o usuário deseja identificar espacialmente a localização de imóveis com mais de 10.000 m² de área de terreno em um determinado distrito, a omissão da informação sobre o centroide da quadra, na qual está localizada o imóvel, leva à omissão da plotagem do referido imóvel e a sua respectiva desconsideração.

4.5.1 Conversão do padrão da base de dados

Os dados relativos ao Cadastro Imobiliário, à tabela de campos e aos valores dos mesmos campos foram fornecidos pela Célula de Tecnologia da Informação (CElTI) da Secretaria de Finanças do Município de Fortaleza (SEFIN), em 07/05/2010, tendo sido gerados em formato texto (TXT), sendo, em seguida, convertidos para o formato *Comma Separated Values* (CSV), separado por vírgulas, com o uso de uma planilha eletrônica. Neste formato, os dados foram lidos pelo *software* aplicativo desenvolvido pelo mestrando, para demonstrar o funcionamento das RNA.

A rotina automática de conversão, utilizada pelo Analisador, segmenta os dados por distrito com o objetivo de otimizar o processamento, uma vez que a Rede Neural é executada sobre um determinado distrito escolhido pelo Operador da Rede. São gerados dois arquivos contendo exclusivamente os campos úteis tratados pelo *software*, sendo desprezados os demais. Ao longo da segmentação são gerados dois arquivos com dados, e três arquivos de relatórios, todos em formato TXT. Um dos arquivos de dados possui linhas de tamanho fixo contendo, em cada uma delas, as características do imóvel à qual se refere, conforme padronização estabelecida pelo programa. Este é o arquivo utilizado nas pesquisas, plotagens e no treinamento da Rede Neural. O outro arquivo é gerado para possibilitar a fácil importação por planilhas eletrônicas, pois atende ao padrão CSV, onde os campos são separados por vírgulas. Já os três relatórios gerados, são os seguintes:

a) **Relatório geral por distrito:** lista a composição de cada distrito em separado, informando a totalização de cada uma das características (atributos) dos imóveis. Por exemplo, a quantidade de barracos, casas, apartamentos, galpões e outros, existentes no distrito.

b) **Relatório geral por atributo:** relaciona os distritos em ordem quantitativa crescente para cada característica, tal como a quantidade de terrenos, de apartamentos, de imóveis não residenciais e outros.

c) **Relatório geral por atributo (planiha):** cria uma consolidação dos quantitativos de cada característica para cada um dos distritos do município, permitindo fácil e rápida comparação de um mesmo atributo entre todos os distritos, através de análise quantitativa ou da geração de estatísticas diversas.

4.5.2 Elaboração das bases de dados auxiliares

A base de dados segmentada é elaborada a partir do Cadastro Imobiliário original, que resta subdividido em 77 arquivos do tipo texto (TXT), denominados "*Quadras do distrito NNN.txt*", onde NNN corresponde à numeração do distrito com três algarismos. Assim, o arquivo gerado para o distrito 15, por exemplo, tem o nome "*Quadras do distrito 015.txt*" e é montado respeitando-se o tamanho fixo definido para cada campo.

Cada linha que compõe o arquivo criado contém as informações sobre as características dos imóveis, desde sua área até a sua classificação arquitetônica, dentre outras, sem qualquer delimitador. Este formato foi utilizado para que, *a posteriori*, pudesse ser trabalhada criptografia específica que garantisse aos dados a necessária segurança à disponibilização e utilização profissionais, fora do escopo acadêmico.

4.5.3 Identificação e dimensionamento dos campos

Os campos lidos pelo Analisador, a partir do banco de dados contendo o Cadastro Imobiliário, foram identificados por arquivo específico de definições, com extensão DEF, fornecido pela SEFIN juntamente com os dados a serem trabalhados. Tais arquivos, também, continham o dimensionamento dos campos, especificando seus nomes, tipos e tamanhos, o que foi utilizado pelo *software* de análise ao tratar da Rede Neural.

Em relação ao Cadastro Imobiliário, foram listados, no Anexo B, os campos utilizados na tributação. Ademais, como no caso do município de Fortaleza, a base de cálculo do IPTU leva em consideração de forma distinta o terreno e a edificação, são relacionados separadamente os campos que influenciam o cálculo do fator de lote, referente à parte territorial, e o do fator de edificação relativo à parte edificada do imóvel.

Vale destacar que, no estudo de caso em apreço, foram utilizadas apenas algumas das características dos imóveis. A seleção abrangeu preferencialmente aquelas características cuja incorreção em seus valores é potencialmente causadora de distorções tributárias. Em estudo futuro mais aprofundado, poder-se-iam incluir outras características igualmente sensíveis e importantes para a tributação.

Vale destacar que outros dados foram anexados ao banco de dados original em virtude de também impactarem na tributação ou de serem resultado de operações realizadas por ocasião do cálculo do imposto. O *fator do lote* e o *fator de edificação*, que originalmente não compõem o banco de dados, foram a ele anexados. Ambos têm por base os dados cadastrais do imóvel, cujo valor venal e valor do IPTU a pagar em 2010 foram igualmente incorporados ao banco de dados, para uso posterior pela Rede Neural.

4.5.4 Análise de dados

Antes que os dados possam ser efetivamente utilizados, o arquivo de dados brutos com 236 MB contendo as informações sobre todo o Cadastro Imobiliário do município de Fortaleza é analisado uma única vez, a fim de otimizar a pesquisa por informações no âmbito de cada distrito, bem como o processamento da própria Rede Neural Artificial. Durante o processamento, quando é feita a segmentação do Cadastro em 77 arquivos menores, correspondendo, um arquivo a cada distrito, inúmeras informações não utilizadas, no presente trabalho, são descartadas, obtendo-se uma redução de espaço de 41,9% em relação ao tamanho do arquivo antes da segmentação.

Todos os arquivos segmentados, juntos, somam 136 MB após a análise e limpeza do banco de dados. Vale salientar que os dados descartados não foram utilizados para a implementação feita nesta dissertação, mas podem vir a ser importantes em estudo futuro que defina nova combinação de características a serem processadas. Os dados foram analisados no equipamento descrito no item 4.2 em aproximadamente 5 minutos, executando as seguintes fases:

- Fase 1: Verificação dos arquivos de leitura.
- Fase 2: Dimensionamento dos campos do cadastro.
- Fase 3: Identificação da dimensão máxima da matriz de campos.
- Fase 4: Dimensionamento da matriz de campos.
- Fase 5: Identificação dos valores possíveis para os campos.
- Fase 6: Análise do cadastro.
- Fase 7: Identificação de quadras por distrito.
- Fase 8: Montagem do relatório.

Ao final da análise, restam criados cinco relatórios, todos em formato texto (TXT) e que são apresentados na sequência:

- a) Relatório GERAL por distrito;
- b) Relatório GERAL por atributo;
- c) Relatório GERAL por atributo (planilha);
- d) Relatório de inconsistências;
- e) Relatório de anormalidades.

4.5.4.1 Relatório GERAL por distrito

O **relatório geral por distrito** é um arquivo único gerado durante a análise do Cadastro Imobiliário e que apresenta quantitativos referentes a algumas características consideradas relevantes para o Operador da Rede Neural ou para a Administração Tributária. Parte do relatório, referente ao distrito 15, é listada a seguir:

Distrito 15

- Total de imóveis	: 10.994
- Quantidade de quadras	: 503
- Quantidade de terrenos	: 63
- Quantidade de edificações	: 10.931
- Quantidade de imóveis residenciais	: 5.931
- Quantidade de imóveis não residenciais	: 4.994
- Quantidade de imóveis públicos	: 65
- Quantidade de imóveis privados	: 10.925
- Quantidade de imóveis imunes	: 210
- Quantidade de apartamentos	: 5.171
- Quantidade de apartamentos de cobertura	: 71
- Quantidade de arquiteturas especiais	: 300
- Quantidade de barracos	: 0
- Quantidade de casas	: 1.275
- Quantidade de conjunto de salas	: 703
- Quantidade de estacionamentos cobertos	: 818
- Quantidade de galpões abertos	: 43
- Quantidade de galpões fechados	: 30
- Quantidade de lojas	: 935
- Quantidade de salas	: 1.350
- Quantidade de sobrelojas	: 2
- Quantidade de subsolos	: 233
- Imóveis com situação normal	: 10.973
- Imóveis com situação bloqueada	: 21
- Imóveis com endereço incompleto	: 0

A diferença encontrada no relatório, entre o total de imóveis e a quantidade de imóveis com situação normal, mostra que existem 21 imóveis em situação anormal no distrito 15. Tais imóveis estão bloqueados por alguma necessidade específica da Administração Tributária, não significando, com isso, que deixem de ser considerados, ficando apenas sob observação especial da Administração.

4.5.4.2 Relatório GERAL por atributo

A geração do **relatório geral por atributo**, que ocorre durante a análise do Cadastro, é de fundamental importância, a fim de permitir a rápida e fácil visualização da composição dos distritos em relação a um determinado atributo, tal como a quantidade de imóveis com apartamentos de cobertura. No relatório são também apresentados os quantitativos de imóveis e quadras em cada distrito, sendo fornecidas, ao todo, as 22 listagens a seguir:

- a) Quanto ao total de imóveis;
- b) Quanto ao total de quadras;
- c) Quanto ao total de terrenos;
- d) Quanto ao total de edificações;
- e) Quanto ao total de imóveis residenciais;
- f) Quanto ao total de imóveis não-residenciais;
- g) Quanto ao total de imóveis públicos;
- h) Quanto ao total de imóveis privados;
- i) Quanto ao total de imóveis imunes;
- j) Quanto ao total de apartamentos;
- k) Quanto ao total de apartamentos de cobertura;
- l) Quanto ao total de imóveis com arquitetura especial;
- m) Quanto ao total de choças ou barracos;
- n) Quanto ao total de casas;
- o) Quanto ao total de conjuntos de salas;
- p) Quanto ao total de estacionamentos cobertos;
- q) Quanto ao total de galpões abertos;
- r) Quanto ao total de galpões fechados;
- s) Quanto ao total de lojas;
- t) Quanto ao total de salas;
- u) Quanto ao total de sobrelojas;
- v) Quanto ao total de subsolos.

A seguir é apresentada a listagem em relação à quantidade de imóveis com apartamentos de cobertura nos distritos, ilustrada na Figura 19. Observa-se que o distrito 63, correspondente aos bairros Barroso e Cajazeiras, localizados na

periferia da capital cearense, aparece como recordista em apartamentos de cobertura, o que é certamente um equívoco. Ocorre que consta no Cadastro Imobiliário, todo um prédio residencial com os 86 apartamentos cadastrados como sendo de cobertura, daí o porquê da incoerência na listagem, na qual o distrito 15, com muita propriedade, aparece trazendo o bairro da Aldeota, em segundo lugar, com 71 apartamentos de cobertura.

Quanto ao TOTAL DE APARTAMENTOS DE COBERTURA

Distrito 021 : 1 (0,14%)	Distrito 013 : 6 (0,84%)
Distrito 022 : 1 (0,14%)	Distrito 008 : 7 (0,97%)
Distrito 025 : 1 (0,14%)	Distrito 017 : 8 (1,11%)
Distrito 037 : 1 (0,14%)	Distrito 027 : 9 (1,25%)
Distrito 040 : 1 (0,14%)	Distrito 012 : 10 (1,39%)
Distrito 045 : 1 (0,14%)	Distrito 053 : 12 (1,67%)
Distrito 046 : 1 (0,14%)	Distrito 077 : 12 (1,67%)
Distrito 050 : 1 (0,14%)	Distrito 056 : 18 (2,51%)
Distrito 062 : 1 (0,14%)	Distrito 052 : 20 (2,79%)
Distrito 068 : 1 (0,14%)	Distrito 026 : 21 (2,92%)
Distrito 071 : 1 (0,14%)	Distrito 016 : 24 (3,34%)
Distrito 076 : 1 (0,14%)	Distrito 011 : 26 (3,62%)
Distrito 006 : 2 (0,28%)	Distrito 005 : 27 (3,76%)
Distrito 010 : 2 (0,28%)	Distrito 014 : 33 (4,60%)
Distrito 039 : 2 (0,28%)	Distrito 018 : 33 (4,60%)
Distrito 043 : 2 (0,28%)	Distrito 051 : 33 (4,60%)
Distrito 049 : 2 (0,28%)	Distrito 048 : 38 (5,29%)
Distrito 004 : 3 (0,42%)	Distrito 047 : 43 (5,99%)
Distrito 007 : 3 (0,42%)	Distrito 009 : 60 (8,36%)
Distrito 036 : 3 (0,42%)	Distrito 019 : 69 (9,61%)
Distrito 038 : 3 (0,42%)	Distrito 015 : 71 (9,89%)
Distrito 055 : 3 (0,42%)	Distrito 063 : 89 (12,40%)
Distrito 001 : 6 (0,84%)	
Distrito 003 : 6 (0,84%)	Total : 718

4.5.4.3 Relatório GERAL por atributo (planilha)

O **relatório geral por atributo (planilha)** é uma variação do **relatório geral por atributo** e traz, em formato CSV, separado por vírgulas, um resumo das totalizações dos campos listados neste último relatório, que já foi comentado anteriormente no tópico 4.5.4.2. O objetivo é permitir fácil acesso a um conjunto de informações consolidadas, que podem ser importadas por planilha eletrônica, para análise de dados posterior. Este relatório contém, em cada linha, o valor de um mesmo atributo para todos os 77 distritos.

Apartamentos de cobertura por distrito

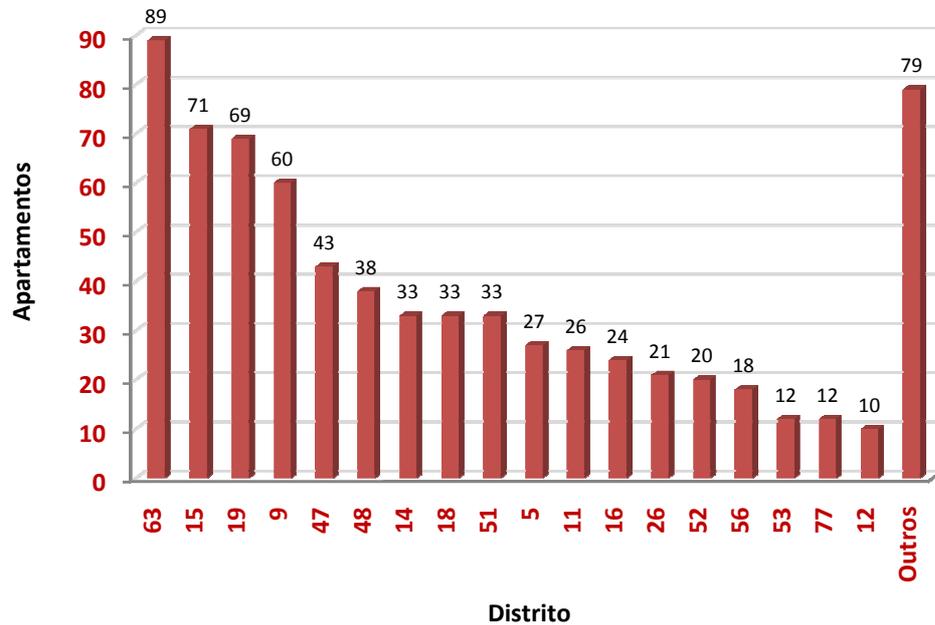


Figura 19 – Imóveis com apartamentos de cobertura nos distritos

4.5.4.4 Relatório de inconsistências

No relatório de inconsistências são listados alguns casos específicos, entendidos como não apropriados para figurar no Cadastro Imobiliário, ou, pelo menos, suspeitos de conter alguma incorreção. Como a Administração Tributária já possui uma série de procedimentos preventivos e corretivos para tratar tais inconsistências, e a sua identificação não é o foco principal deste trabalho, o relatório se limita a ilustrar que a análise dos dados também pode ser utilizada para criticar mais detalhadamente as características cadastrais dos imóveis, bastando que se adapte o *software*, para realizar outras críticas e cruzamentos de dados.

O Quadro 5 indica, em ordem alfabética, os tipos de ocorrências verificadas, bem como os quantitativos encontrados para cada uma delas durante a rotina de análise, destacando que, um mesmo imóvel, pode aparecer em mais de uma ocorrência distinta, caso sejam identificadas, para ele, duas ou mais inconsistências ou suspeições.

Inconsistência ou suspeição	Ocorrências
Imóvel não territorial sem área edificada	632
Imóvel predial com mais de 30 pavimentos	1.616
Imóvel predial sem número de unidades	92.934
Imóvel público, não imune	604
Imóvel territorial com classificação arquitetônica de casa ou outra	3.692
Imóvel territorial contendo área edificada	3.335
Quantidade de unidades muito grande (> 300)	16.518
Testada muito grande (> 200 m)	3.697
Valor venal maior que R\$ 10 milhões	106
Valor venal menor que R\$ 500,00	69.263

Quadro 5 – Inconsistências ou suspeições detectadas durante a análise do Cadastro

4.5.4.5 Relatório de anormalidades

O relatório de anormalidades lista os imóveis cujo cadastro foi excluído, pelos mais diversos motivos, tal como ilustra o Quadro 6. Seu objetivo é proporcionar uma simples listagem, para conferência, uma vez que a exclusão inapropriada de um imóvel do cadastro Imobiliário, causa perda de receita tributária para o Município, ainda com inúmeros impactos administrativos, para não mencionar os desdobramentos civis e penais decorrentes de sua exclusão indevida, como citam Pinto, Windt e Céspedes (2004, p.131) ao lembrar que, a inserção de dados falsos em sistemas de informações da Administração Pública, constitui crime contra ela.

4.6 Definição da pesquisa

A pesquisa foi definida para ter, como escopo, o distrito 15 do município de Fortaleza, correspondente ao bairro da Aldeota. Uma vez que o objetivo da pesquisa é a identificação de distorções tributárias, mesmo dados inconsistentes foram conservados, a fim de que também pudessem ser classificados de forma adequada, já que possuem impacto tributário.

Antes que efetivamente possa ser iniciada, a pesquisa necessita que sejam definidos dois elementos fundamentais, a saber: a seleção da área exploratória e a seleção dos parâmetros de pesquisa.

Anormalidade	Ocorrências
Cancelado (cód.2)	9718
Demolido (cód.5)	270
DESCONHECIDO (cód.0)	2
DESCONHECIDO (cód.16)	1
DESCONHECIDO (cód.26)	1
DESCONHECIDO (cód.31)	3
Domínio público (cód.4)	2066
Duplicidade (cód.7)	6780
Fenômeno físico (cód.6)	20
Fora do limite (cód.8)	4211
Remembrado (cód.3)	25773

Quadro 6 – Anormalidades causadoras de exclusão de imóveis do cadastro Imobiliário

4.6.1 Seleção da área exploratória

Em virtude da grande extensão territorial do município e, especialmente, da grande quantidade de imóveis no Cadastro Imobiliário, resolveu-se limitar a pesquisa a uma área específica da cidade de Fortaleza, uma vez que, mesmo com esta limitação, a quantidade e a variedade de dados não foram prejudicadas, também não redundando em qualquer prejuízo para a demonstração do funcionamento das Redes Neurais Artificiais.

A opção pelo distrito 15, correspondente ao bairro Aldeota, deveu-se à necessidade de pôr à prova o mecanismo de identificação de distorções tributárias justamente em uma das áreas onde se considera que há maior atualização cadastral. É nesta amostra, então, que a Rede Neural apresenta sua habilidade para identificar distorções além do que já foi exaustivamente percebido, acompanhado e atualizado pelo Fisco.

A análise das características do distrito 15, quanto à sua composição imobiliária, encontra-se descrita na Tabela 5 que traz, também, os valores mínimo e máximo encontrados no levantamento feito com todos os 77 distritos do município de Fortaleza em relação a itens como a quantidade de terrenos ou a quantidade de imóveis residenciais identificados. Observa-se, por exemplo, que o município possui 561.341 edificações em seu Cadastro Imobiliário, sendo que a menor quantidade de edificações, 196 inscrições apenas, é encontrada no distrito 58.

Já a maior quantidade de edificações, ou seja, 14.730 inscrições, está no distrito 19. A média de edificações por distrito é de 7.290. No distrito 15, verifica-se a existência de 10.931 edificações. No Anexo B, são listados os nomes dos bairros abrangidos por cada distrito, total ou parcialmente, valendo observar que os limites dos distritos usualmente não coincidem com os limites dos bairros. Os dados apresentados foram fornecidos pela Secretaria de Finanças do Município de Fortaleza, e correspondem à situação do Cadastro Imobiliário em 07/05/2010.

4.6.2 Seleção dos parâmetros de pesquisa

O município de Fortaleza possui, em sua legislação, a indicação precisa da forma de se calcular o IPTU de cada imóvel, tanto no que concerne à base de cálculo quanto à alíquota. Os anexos II e IV da lei 8703 de 30/04/2003, que definem a fórmula de cálculo do IPTU ainda vigente no Município de Fortaleza e os valores em reais para o metro quadrado construído, conforme as características da edificação, são apresentados nos Anexos C e D. Convém lembrar que os valores constantes do Anexo D vêm sendo atualizados desde 2004 pela inflação acumulada em cada ano fiscal, tendo havido atualização diferenciada em 2009, a vigor no exercício de 2010, cujo comentário será abstraído, por não constituir objetivo deste trabalho. Frise-se, contudo, que os dados utilizados nos estudos apresentados já se encontram devidamente atualizados.

Tecnicamente, a base de cálculo, ou seja, o valor venal estimado para cada imóvel, toma para a sua formação diversos elementos que podem ser divididos em três grupos principais:

- a) os elementos relativos à caracterização do terreno;
- b) os elementos relativos à caracterização da edificação;

- c) as medidas lineares (m) e de área (m²);
- d) os valores definidos em lei para o m² do terreno de acordo com a face de quadra em que está localizado o imóvel, e o valor do m² da edificação para o padrão construtivo identificado a partir dos materiais utilizados; e
- e) fatores de proporcionalidade para o terreno (fração ideal) e para a edificação (depreciação).

Os elementos relativos ao terreno indicam características do lote de terra no qual está inserido o imóvel ou que com ele coincide, caso o imóvel não seja edificado, tais como, por exemplo: a situação do lote (encravado, normal, esquina, quadra, gleba...), o tipo de calçada (cimentada, com pré-moldado, mosaico...), a pavimentação para veículos (paralelepípedo, pré-moldado, asfalto, concreto...), a existência de iluminação pública e de rede de esgoto, dentre outros.

Os elementos relativos à edificação indicam, por exemplo, a classificação arquitetônica do imóvel (barraco, casa, apartamento, sala, galpão, estacionamento...), equipamentos como piscina e jardim, além do próprio estado de conservação do imóvel. Também fazem referência ao material utilizado no piso (cerâmica, granito, mármore...) e nos revestimentos (cal, pintura impermeável, textura...), elementos que permitem inferir o nível de qualidade e, conseqüentemente, de valor a partir do tipo de acabamento do imóvel.

As medidas lineares e de área dizem respeito às dimensões do imóvel, seja quanto ao terreno, seja quanto à edificação. Englobam três dados distintos e igualmente relevantes, que são: a **área do terreno**, a **testada principal** ou seja, a extensão em metros da frente do terreno, e a **área edificada** do imóvel.

O Cadastro Imobiliário é composto por uma vasta quantidade de dados relativos a cada imóvel. Para a abordagem acadêmica à qual esta dissertação se propõe, foi suficiente a aplicação da Rede Neural de Kohonen a apenas uma parte destes dados. Foram escolhidas, então, as características dos imóveis cujas incorreções são passíveis de gerar maior perda de receita aos cofres municipais, evasão esta não identificável pelos mecanismos tradicionais de extração e pesquisa. Corrigi-las traz especial impacto positivo para a tributação e sua mera identificação torna a fiscalização muito mais eficaz.

Ao conjunto de características listadas a seguir, e selecionadas do banco de dados imobiliário para utilização ao longo deste estudo, através da aplicação das

Redes Neurais de Kohonen, dá-se o nome de **parâmetros de pesquisa**, valendo salientar que qualquer combinação de características poderia ter sido selecionada, a partir do entendimento do conhecedor das regras de cada negócio:

- a) a área do terreno (em m²);
- b) a área edificada (em m²);
- c) a classificação arquitetônica;
- d) o tipo do imóvel;
- e) o uso específico; e
- f) o valor do IPTU no exercício de 2010 (em R\$).

4.6.3 Pesquisa da composição imobiliária

A pesquisa da composição imobiliária pode ser realizada em todo o município, ou em apenas um de seus distritos. Tem como objetivo a identificação e localização de imóveis com uma característica específica, como, por exemplo, área territorial acima de 1.000 m².

A Figura 20 ilustra o ambiente de seleção, através do qual, o usuário escolhe o distrito e o respectivo atributo a ser pesquisado, dentre outros parâmetros. Ao serem processados os imóveis, pertencentes ao distrito selecionado, são computadas as ocorrências do atributo escolhido, dentro do limite especificado, em cada quadra, e o resultado é mostrado, tanto por ordem crescente dos números das quadras, quanto por ordem decrescente da quantidade de ocorrências, tal como pode ser observado na listagem do Apêndice A.

Já a Figura 21, exemplifica como são demonstradas a localização e as ocorrências referentes aos imóveis, que possuem o atributo, ou característica, selecionada para pesquisa. No exemplo apresentado, a característica escolhida foi a área do terreno, sendo ainda indicado que se desejava pesquisar somente os imóveis, do distrito 15, com áreas maiores que 1.000 m².

Tabela 5 – Comparação quantitativa do distrito 15 com o município

	Total no município	Mínimo em um distrito	Distrito	Máximo em um distrito	Distrito	Média nos distritos	Distrito 15 (Aldeota)
IMÓVEIS	656.571	1.622	54	25.119	71	8.527	10.994
QUADRAS	20.579	9	67	1.491	19	267	503
TERRENOS	95.201	16	67	11.594	71	1.236	63
EDIFICAÇÕES	561.341	196	58	14.730	19	7.290	10.931
IMÓVEIS RESIDENCIAIS	444.741	154	58	12.668	68	5.776	5.931
IMÓVEIS NÃO RESIDENCIAIS	115.877	42	58	6.780	19	1.505	4.994
IMÓVEIS PÚBLICOS	4.568	3	77	346	68	59	65
IMÓVEIS PRIVADOS	651.387	1.609	54	25.022	71	8.460	10.925
IMÓVEIS IMUNES	16.342	11	58	1.398	51	212	210
APARTAMENTOS	135.054	0	58	9.605	9	1.754	5.171
APARTAMENTOS DE COBERTURA	718	0	2	89	63	9	71
IMÓVEIS COM ARQUITETURA ESPECIAL	2.599	1	55	300	15	34	300
CHOÇAS OU BARRACOS	16.437	0	2	918	68	213	0
CASAS	330.190	57	54	11.496	29	4.288	1.275
CONJUNTOS DE SALAS	6.508	2	58	703	15	85	703
ESTACIONAMENTOS COBERTOS	4.636	0	23	1.483	19	60	818
GALPÕES ABERTOS	2.728	0	67	98	62	35	43
GALPÕES FECHADOS	5.665	5	55	298	62	74	30
LOJAS	35.667	11	58	2.218	2	463	935
SALAS	22.824	3	54	2.008	1	296	1.350
SOBRELOJAS	82	0	3	27	19	1	2
SUBSOLOS	1.298	0	2	443	9	17	233

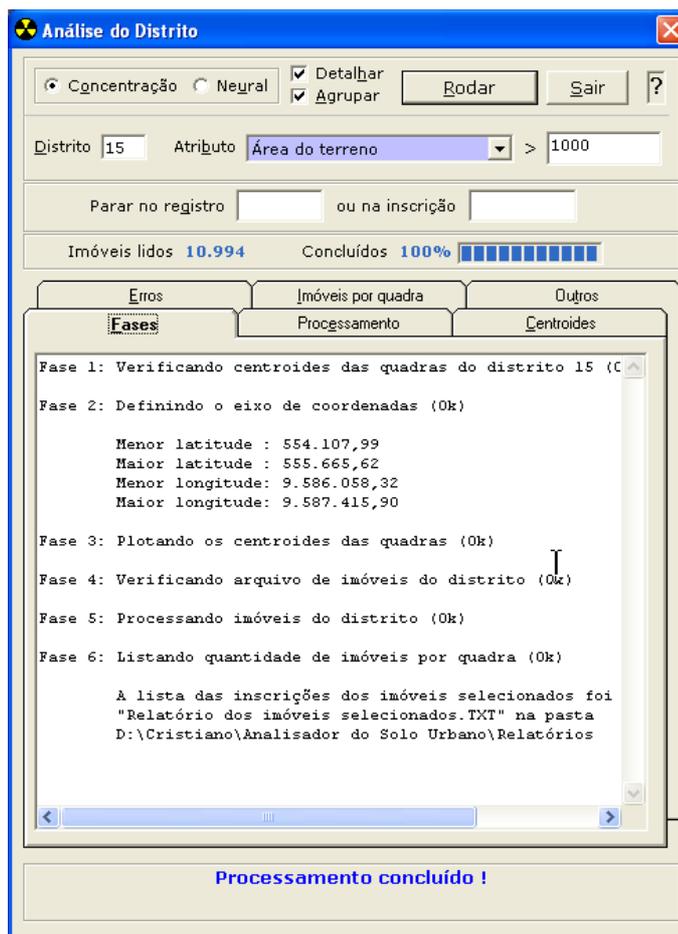


Figura 20 – Pesquisa de composição imobiliária através da análise do distrito

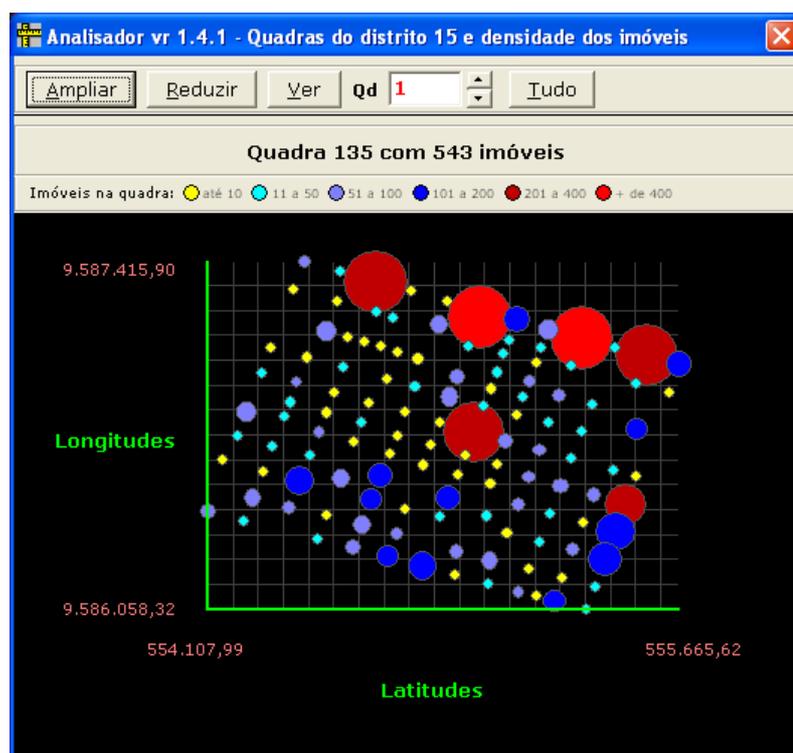


Figura 21 – Ocorrências de imóveis com áreas de terreno maiores que 1.000 m² no distrito 15



Figura 22 – Legenda sobre a quantidade de imóveis identificados na quadra

O posicionamento dos imóveis, no plano cartesiano representativo das respectivas localizações geográficas, é feito pelo centroide da quadra onde estão localizados. De acordo com a quantidade de ocorrências encontradas, em uma determinada quadra, para a característica selecionada pelo usuário, são desenhados círculos de cores e tamanhos distintos. As cores variam de acordo com faixas de quantidades previamente estabelecidas, ilustradas na Figura 22, a fim de facilitar a visualização, pelo usuário, das quadras com quantidades semelhantes de imóveis. Os tamanhos dos círculos são proporcionais às quantidades de ocorrências em cada quadra, sendo possível, no Analisador, navegar pelo ambiente e identificar, passando o cursor sobre o círculo correspondente à quadra desejada, a quantidade de ocorrências identificadas na quadra em questão.

4.7 Experimento de treinamento e classificação

Uma vez indicado o distrito sobre o qual se deseja executar a Rede de Kohonen, passa-se, então, a preparar o treinamento, especificando-se, basicamente, as características que irão compor a Camada de Entrada, cujos dados serão analisados, e os critérios de parada.

4.7.1 Configuração da Camada de Entrada

A Camada de Entrada contém o indicativo das características dos imóveis, que serão analisadas. Cada característica equivale a um neurônio na Camada de Entrada, e cada um destes neurônios tem um peso representativo da força de sua ligação com os neurônios da Camada de Saída. Aconselha-se uma revisita à Figura 5 na página 47 para melhor assimilação a respeito da forma como ocorrem as ligações entre ambas as camadas.

Camada de entrada		
Menor valor	NEURÔNIO	Maior valor
1,00	Área do terreno (m ²)	33.898,00
0,00	Área edificada (m ²)	15.000,00
0	Classificação arquitetônica	13
0	Nº de pavimentos	43
0	Nº de unidades	503
1	Ocupação	7
1	Patrimônio	5
1	Posição fiscal	6
0,00	Testada (m)	9.500,00
1	Tipo do imóvel	2
0	Uso específico	16
0,00	Valor venal (R\$)	34.927.039,00
0,00	Valor do IPTU (R\$)	696.434,00
0	Fator do lote	1
0	Fator de edificação	1

Figura 23 – Seleção de neurônios para a composição da Camada de Entrada

Na Figura 23 pode ser visualizado o módulo de seleção dos neurônios, ou seja, das características dos imóveis, que vão compor as entradas durante o treinamento. Estando identificados, em virtude de processamento anterior, o menor e o maior valor possível para cada neurônio selecionado, tal como se depreende da mesma figura, a definição das classes e de seus limites torna-se viável para fins de normalização.

O *software* foi implementado de forma a marcar, através de cor diferenciada, os neurônios de entrada selecionados para treinamento, o que pode ser realizado com um simples clique do mouse sobre o nome do neurônio desejado. Para desmarcar o neurônio, basta um novo clique sobre o nome pelo qual é identificado.

4.7.2 Configuração da Camada de Saída

O nível de especialização, isto é, de precisão, da Rede Neural, dependerá da quantidade de neurônios na Camada de Saída. É nesta camada que acontece a competição entre todos os seus neurônios, a fim de se consagrar vencedor, aquele que possui maior afinidade com a entrada apresentada à RNA.

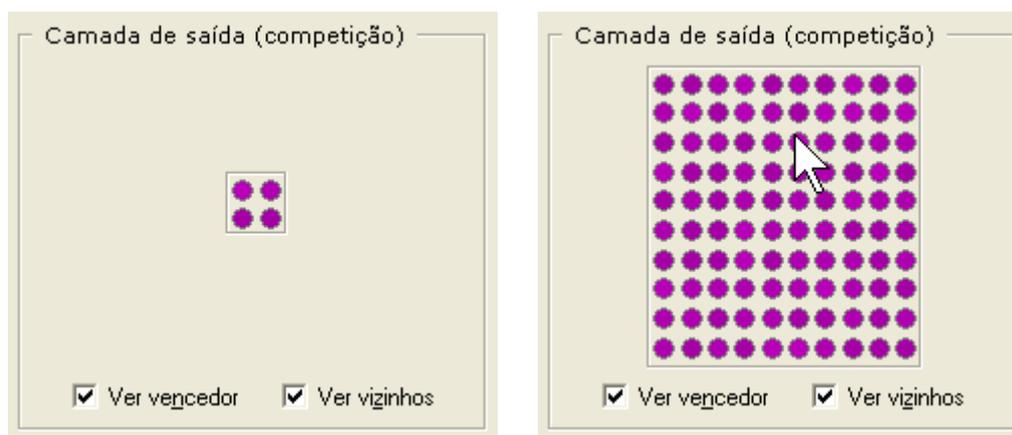


Figura 24 – Camada de saída, também chamada de Camada de Competição

Em nosso estudo, esta Camada é definida com 100 neurônios, em uma matriz 10x10. No entanto, o *software* Analisador do Solo Urbano em sua versão 1.4.1 permite qualquer outra dimensão $N \times N$, com N variando de 2 a 10. A definição pode ser feita em qualquer momento, desde que, anteceda o início do treinamento da rede. Ao final do aprendizado, cada neurônio da Camada de Saída mapeará um grupo de entradas com características semelhantes, a partir do aprendizado adquirido pela RNA.

Na Figura 24 podem ser visualizadas duas configurações da Camada de Saída. A configuração, mais à esquerda, representa uma camada com apenas 4 neurônios, com dimensão quadrada, 2x2, possuindo pequena especialização, uma vez que existem poucas ligações com os neurônios de entrada e,

consequentemente, limitadas possibilidades para ajustes de pesos. Na configuração mais à direita, a camada apresenta 100 neurônios dispostos em uma matriz 10x10. Justamente em virtude da grande quantidade de saídas, que cria uma vasta teia de ligações entre elas e as entradas apresentadas, o treinamento alcança grande especialização, favorecendo a classificação adequada das entradas, pela Rede Neural Artificial.

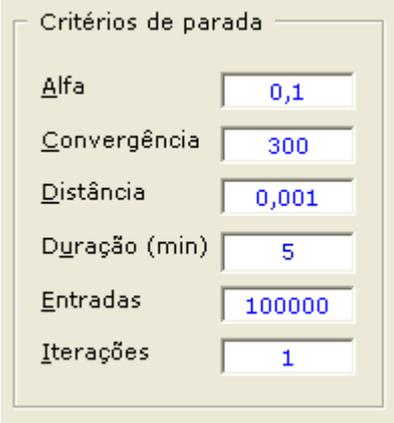
4.7.3 Definição dos critérios de parada

Os critérios de parada são parâmetros indicativos de que o aprendizado já alcançou níveis satisfatórios, em decorrência do que, deve ser terminado o treinamento. Não há limites, para a quantidade de critérios de parada, podendo ser criados tantos quantos se deseje, para o melhor desempenho da rede. A partir do maior e melhor conhecimento da metodologia de trabalho das RNA, novos e complexos modelos podem ser definidos, valendo lembrar que todos devem ser continuamente testados para que, ao serem alcançados, o treinamento seja interrompido.

Os critérios de parada adotados pelo *software*, para processar o treinamento não supervisionado, implementado com o objetivo de demonstrar o funcionamento e o uso da Rede Neural de Kohonen, são os seguintes e que também se encontram ilustrados na Figura 25:

- a) **Alfa (α)**: também chamado de **taxa de aprendizagem**, é um multiplicador que desacelera o treinamento, dando-lhe maior capacidade de especialização. Alfa dá relevância ao ajuste de pesos. O valor de alfa é definido entre zero e um, sendo inicializado normalmente com valores altos, tais como 0,9, antes que se inicie o treinamento. Na medida em que o treinamento evolui, o valor de alfa é decrementado. Quando chega a valores próximos de zero, como 0,1, por exemplo, a rede passa a adquirir novo aprendizado de forma muito lenta, favorecendo-se extremamente a especialização e caindo drasticamente o nível de aprendizagem (β), definido pelo mestrando como $1-\alpha$. A diminuição de alfa com o tempo deve ser ajustada adequadamente. Na versão atual, o *software* define que a taxa de aprendizagem, no início do treinamento, tem o valor 0,9, sendo diminuída

em 1% a cada 10 segundos, em relação ao seu valor imediatamente anterior. A Figura 26 ilustra como o nível de aprendizagem tende a ficar inexpressivo com o tempo.



Critérios de parada	
Alfa	0,1
Convergência	300
Distância	0,001
Duração (min)	5
Entradas	100000
Iterações	1

Figura 25 – Critérios de parada adotados pelo Analisador

- b) **Convergência**: indica a quantidade de vezes que o mesmo neurônio da Camada de Competição se consagra vencedor. Ora, se o mesmo neurônio sempre se consagra vencedor, não há mais razão para ajustes nos pesos e, portanto, não persiste o aprendizado, podendo, a rede, ser considerada treinada.
- c) **Distância**: quando a distância entre as entradas apresentadas e os neurônios da Camada de Saída são muito pequenas, então já existe grande concentração de neurônios nos grupos. Assim, a rede já pode ser considerada treinada, pois não há tendência à dispersão dos grupos formados.
- d) **Duração**: a duração máxima, em minutos, permitida para o treinamento, pode ser configurada por razões elementares. Se assim não fosse, um determinado treinamento que não alcançasse nenhum outro critério de parada persistiria por horas, dias, meses, enfim. Daí, dentro de limites entendidos como razoáveis, o tempo máximo de treinamento pode e deve ser estabelecido, conforme seja julgado ideal, pelo Operador da Rede. Quando estiver disponível grande quantidade de dados, por exemplo, pode-se estabelecer como limite um decurso de prazo que se considere suficiente para que a rede alcance razoável nível de especialização. Senão, vejamos: a simulação de uma estrutura complexa contendo 15 neurônios na Camada de Entrada, 100 neurônios na Camada de Saída e 500 mil dados, exigiu 11

minutos de pré-processamento e outros 89 minutos para completar um único ciclo de treinamento no equipamento descrito no Quadro 3. O tempo pode ser julgado demasiadamente elástico, especialmente em virtude de ainda ser necessário igual prazo para a classificação dos dados.

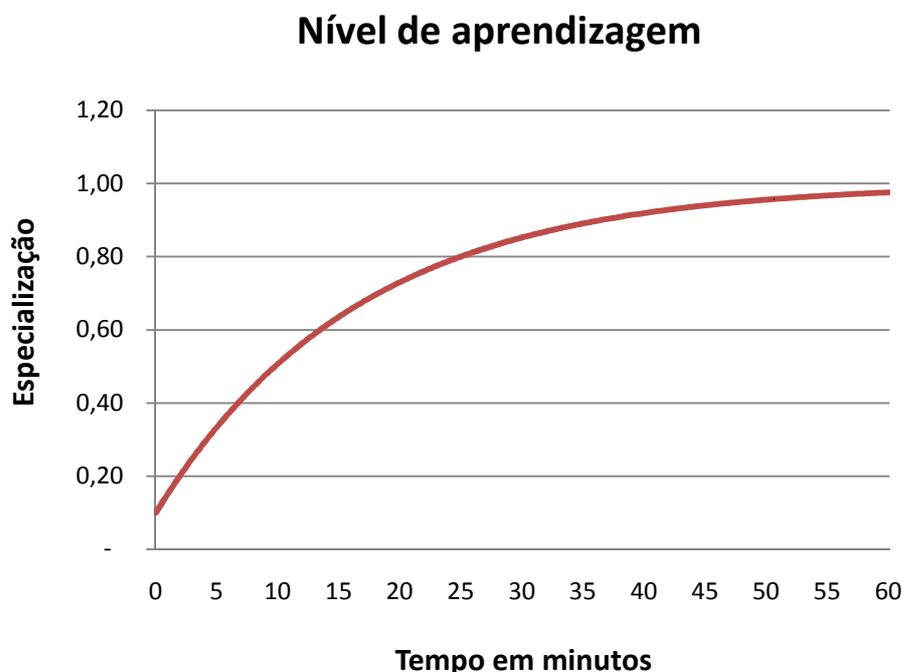


Figura 26 – Desaceleração na aprendizagem com o tempo

- e) **Entradas**: a quantidade máxima de entradas que se deseja processar, pode ser expressamente declarada como critério de parada. Atingida esta quantidade, a rede para, pois é considerada treinada. Uma pequena quantidade de entradas pode ser insuficiente para um bom treinamento. Uma quantidade de entradas excessiva pode levar a rede ao supertreinamento, fenômeno que ocorre, quando a rede fica “acostumada” com os dados de entrada, tornando-se incapaz de classificar corretamente outros dados que não aqueles anteriormente utilizados no treinamento. Quando o Operador da Rede desejar processar um número de entradas equivalente à totalidade de dados disponíveis, basta ajustar o número de iterações para um. Em contrapartida, usualmente em virtude de estar disponível para o treinamento uma grande quantidade de dados, o número de entradas pode ser limitado de acordo com a quantidade que se julgue ser representativa de todos os agrupamentos distintos a serem formados.

f) **Iterações:** denotam a quantidade de vezes que o conjunto de entradas é processado. Vale destacar que, ao contrário do que se poderia imaginar, uma nova iteração, ofertando novamente à rede o mesmo conjunto de dados para treinamento, leva a aprendizado diferenciado do ocorrido na iteração, ou ciclo, anterior. Isto porque os dados apresentados à rede por meio de um mesmo neurônio de entrada, não necessariamente têm contribuição única e constante para o treinamento, haja vista que os neurônios da Camada de Saída competem em relação à entrada apresentada, de acordo com os pesos que possuem no momento da apresentação da entrada à rede. Como os pesos são dinâmicos, o efeito produzido por uma mesma entrada em momentos distintos, também é dinâmico. Quando os dados são escassos, após o processamento de todo um ciclo, ou seja, quando a rede é treinada a partir da análise de cada um deles, é possível submetê-la novamente ao mesmo conjunto de dados. Como a similaridade já foi analisada na passagem do primeiro ciclo, o processamento repetido de todo o conjunto de dados, em uma nova iteração, pode vir a dar maior nível de especialização à RNA. Uma única iteração pode ser suficiente para que a rede obtenha um nível aceitável de conhecimento e especialização.

Conclui-se, portanto, que, nem sempre o treinamento realizado durante um maior tempo, ou considerando uma maior quantidade de entradas, é o que gera melhores resultados no momento da classificação. De forma análoga, o estabelecimento de critérios de parada precisa ser suficiente e exaustivamente testado, para que se chegue aos melhores pesos e, daí, à melhor e mais precisa classificação.

A configuração de todos os critérios de parada, a serem utilizados no treinamento da Rede de Kohonen, pode ser feita no Analisador, em módulo específico, apresentado na Figura 27, lembrando que, podem ainda ser indicados quais os critérios de parada que devem ser utilizados, e quais os que devem ser relevados.

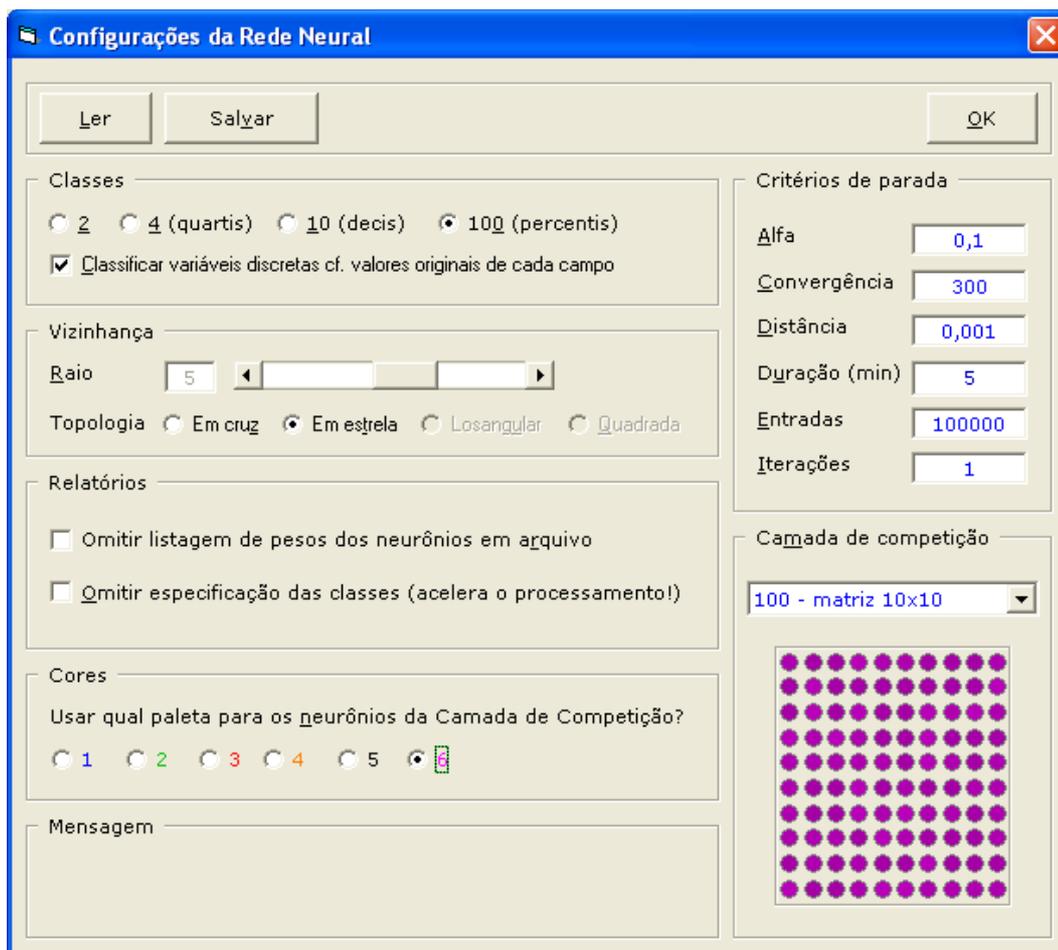


Figura 27 – Configuração dos parâmetros de treinamento no Analisador

4.7.4 Normalização dos dados

Ao serem apresentadas à rede para participarem do treinamento, as entradas são normalizadas, isto é, codificadas em intervalo de grandeza única, normalmente entre 0 e 1, a fim de que tenham tratamento homogêneo durante o processamento da RNA, tal como já foi amplamente explicado no tópico 3.5.

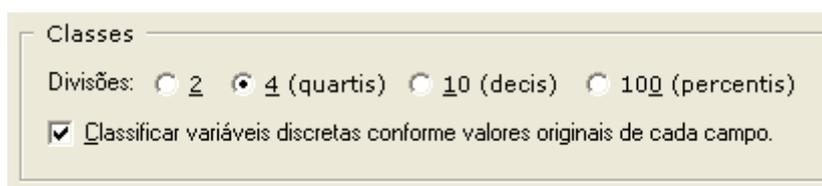


Figura 28 – Definição de classes para a normalização das entradas

A normalização é feita enquadrando-se a entrada em um valor representativo de uma das classes, que são definidas a partir da quantidade de

subdivisões escolhida pelo Operador da Rede. A Figura 28 ilustra como o Operador indica:

a) a quantidade de classes, em que deseja ver os dados relativos a variáveis contínuas, subdivididos;

b) a forma de subdivisão dos dados relativos a variáveis discretas: usando a mesma quantidade de classes definida para as variáveis contínuas, ou usando tantas classes quantos forem os dados possíveis para a variável discreta em questão.

A seguir, são apresentados os resultados da normalização, para dois neurônios de entrada: o *patrimônio* e a *área edificada*. Vale lembrar que, antes do treinamento, os valores mínimo e máximo possíveis, para ambos os neurônios, já haviam sido definidos pelo *software*.

NEURÔNIOS DA CAMADA DE ENTRADA

1. Área edificada (m²)
2. Patrimônio

QUANTIDADE DE CLASSES: 4 (variáveis discretas classificadas conforme valores originais dos campos)

DEFINIÇÃO DOS LIMITES DAS CLASSES

Neurônio de entrada: "Área edificada (m²)"

- Menor valor.....: 0,00
- Maior valor.....: 15.000,00
- Amplitude: 15.000,00

- Quantidade de classes: 4
- Intervalo de classe..: 3.750,00

Classe 1

- Limite inferior : 0,00
- Limite superior : 3.750,00
- Valor normalizado: 0,25

Classe 2

- Limite inferior : 3.750,00
- Limite superior : 7.500,00
- Valor normalizado: 0,50

Classe 3

- Limite inferior : 7.500,00
- Limite superior : 11.250,00
- Valor normalizado: 0,75

Classe 4

- Limite inferior : 11.250,00
- Limite superior : 15.000,00
- Valor normalizado: 1,00

Neurônio de entrada: "Patrimônio"

- Menor valor.....: 1,00
- Maior valor.....: 5,00

- Quantidade de classes: 5

- Classe 1 = "1"
- Classe 2 = "2"
- Classe 3 = "3"
- Classe 4 = "4"
- Classe 5 = "5"

Para efeito de melhor visualização, o Quadro 7 ilustra como ficam os neurônios de entrada após suas respectivas normalizações, caso o Operador da Rede tenha configurado a quantidade de classes das variáveis contínuas para 4. Suponha-se, por exemplo, que uma entrada apresentada à rede traga o valor 4.000,00 para o neurônio *área edificada*. Pelo quadro mencionado, é possível deduzir que o neurônio terá como valor normalizado 0,50, uma vez que este valor está compreendido entre 3.750,00 e 7.500,00.

Neurônio: área edificada			
Classe	Limite inferior	Limite superior	Valor normalizado
1	0,00	3.750,00	0,25
2	3.750,00	7.500,00	0,50
3	7.500,00	11.250,00	0,75
4	11.250,00	15.000,00	1,00

Quadro 7 – Normalização de variável contínua

Já no caso do neurônio patrimônio, por se tratar de uma variável discreta, tem-se a situação ilustrada no Quadro 8, quando cada valor possível para a entrada, dá origem a uma classe, uma vez que o Operador da Rede expressou o seu desejo de “classificar variáveis discretas, conforme o valor original de cada campo”, ao marcar esta opção no *software*.

Neurônio: patrimônio	
Classe	Valor normalizado
1	0,20
2	0,40
3	0,60
4	0,80
5	1,00

Quadro 8 – Normalização de variável discreta

4.7.5 Treinamento da rede

Diante dos parâmetros de pesquisa selecionados tal como descrito no item 4.6.2, a RNA proposta neste trabalho realiza o treinamento respeitando os demais parâmetros especificados pelo Operador da Rede, conforme apresentado no Quadro 9. Ao ser definida a quantidade máxima de iterações como sendo de 1 unidade, significa que a rede será considerada treinada após o primeiro ciclo, ou seja, após processar a primeira vez, todas as entradas, interrompendo o aprendizado e passando à fase de classificação. Vale lembrar que, a parametrização da rede deve ser realizada de forma cautelosa e apropriada, para um melhor aproveitamento do aprendizado e, conseqüentemente, para uma maior adequação da classificação.

Parâmetro	Nome	Valor
1	Classes para variáveis contínuas	100
2	Classes para variáveis discretas	Conforme cada campo
3	Raio da vizinhança	5
4	Topologia da vizinhança	Em estrela
5	Iterações	1

Quadro 9 – Parametrização da Rede Neural para treinamento.

Iniciado o treinamento, cada novo dado é apresentado à rede sob a forma de um neurônio de entrada, sendo então calculadas as distâncias euclidianas entre o neurônio de entrada e cada neurônio da Camada de Saída. A menor distância encontrada determinará o Neurônio Vencedor na Camada de Saída, indicando que aquela entrada tem maior afinidade com o neurônio vencedor, do que com todos os demais neurônios da Camada de Saída, por isso mesmo, mantendo uma ligação, ou peso, mais forte.

Apenas para efeito de facilitação do entendimento, suponha-se uma Rede Neural Artificial contendo como neurônios da Camada de Entrada, a *área edificada* e o *tipo do imóvel*. O treinamento da rede, que ocorre através do ajuste de pesos das ligações entre as entradas e os neurônios da Camada de Saída é explicado, passo-a-passo, no Apêndice B, a partir de parametrização feita pelo Operador da Rede, que definiu, para o caso ilustrado, os seguintes parâmetros:

- a) Neurônios da Camada de Entrada: *área edificada* e *tipo do imóvel*;
- b) Ordem da Camada de Saída: 2x2 (4 neurônios);
- c) Topologia da vizinhança: em estrela; e
- d) Raio da vizinhança: 5 unidades;

4.7.6 Classificação

Classificar os dados consiste em apresentar cada um deles novamente à rede, agora treinada, a fim de que seja identificado com qual dos neurônios, a entrada apresentada possui maior afinidade. O procedimento da classificação é praticamente o mesmo do treinamento, à exceção de que, na classificação, não existe mais ajuste de pesos, nem do neurônio vencedor, nem de seus vizinhos.

Desta forma, apresentada a entrada à RNA, identifica-se o neurônio vencedor e, por fim, a entrada é associada ao grupo formado a partir do neurônio vencedor. Após a apresentação de todas as entradas, poderão ser formados tantos grupos quantos sejam os neurônios da Camada de Saída. Naturalmente, na pior das hipóteses, todos os neurônios de entrada ficarão agrupados juntos, formando, com um dos neurônios da Camada de Saída, um único grupo.

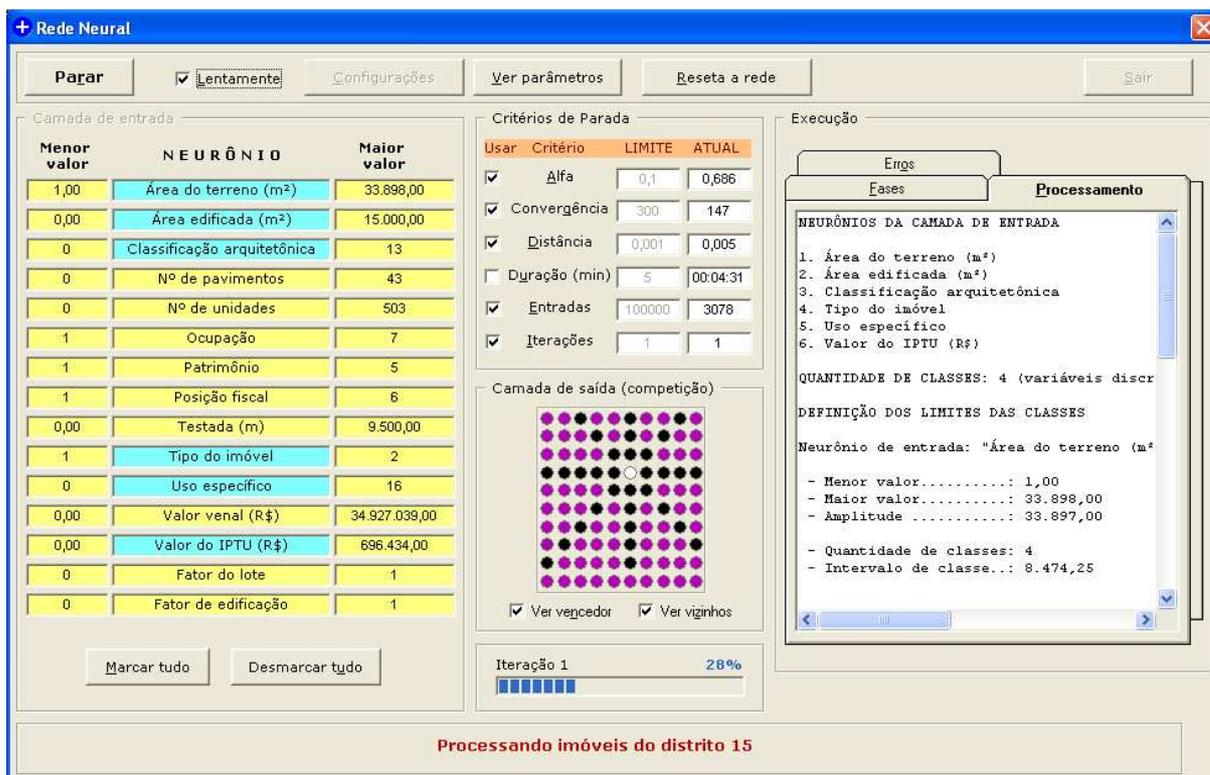


Figura 29 – Ambiente de treinamento da Rede Neural de Kohonen

No Quadro 10, bem como no gráfico representado na Figura 30, são apresentados diversos tempos, obtidos pelo microcomputador mencionado no item 4.2, para o treinamento da rede e classificação dos dados de entrada, cuja amostra é composta de 10.994 imóveis do distrito 15, o bairro da Aldeota na cidade de Fortaleza.

Em resumo, no caso do Cadastro Imobiliário, após classificadas todas as entradas referentes a um mesmo distrito, restam gerados grupos de imóveis com características semelhantes.

Dimensão da Camada de Saída	Neurônios na Camada de Saída	Duração do treinamento	Duração da classificação	Duração total	Agrupamentos formados
2x2	4	10s	18s	28s	3
4x4	16	16s	22s	38s	9
6x6	36	24s	29s	53s	16
8x8	64	34s	38s	1m12s	30
10x10	100	45s	50s	1m35s	46

Quadro 10 – Quadro comparativo de performance para treinamento e classificação.

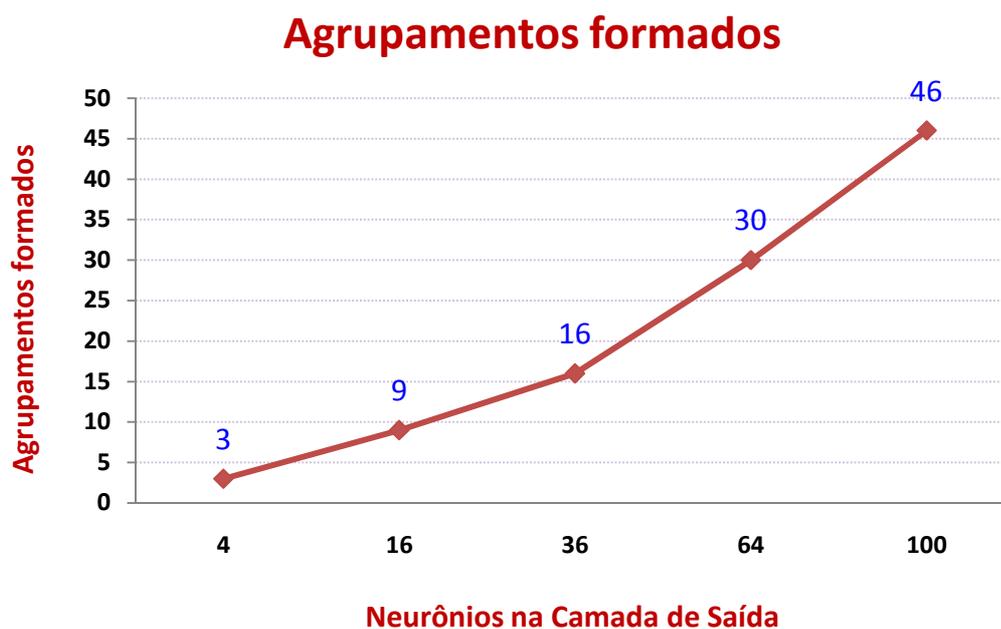


Figura 30 – Relação entre os neurônios de saída e os agrupamentos formados

Observa-se uma maior especialização da RNA na medida em que aumenta o número de neurônios da Camada de Saída, permitindo que haja maior diferenciação entre os agrupamentos formados. A definição da quantidade de neurônios na Camada de Saída é facultada ao Operador da Rede, quem deve ter como base de decisão os tipos de neurônios de entrada a serem classificados, e os valores que tais neurônios podem assumir.

Ao final do treinamento, os pesos definidos para cada neurônio habilitam a rede a identificar, para qualquer entrada que lhe for apresentada, o grupo com o qual a entrada considerada possui maior afinidade.

Daí, apresentadas à rede todas as entradas, estas serão agrupadas de acordo com os pesos definidos, gerando grupos de imóveis com características semelhantes, tais como os anteriormente ilustrados na Figura 18.

O resultado final da classificação das entradas foi a formação de 40 grupos distintos contendo imóveis semelhantes. As quantidades de entradas alocadas em cada grupo estão listadas no Quadro 11, e a distribuição das ocorrências está ilustrada no gráfico da Figura 31.

Neurônio	Imóveis agrupados	Neurônio	Imóveis agrupados	Neurônio	Imóveis agrupados
95	3140	2	202	96	31
84	2030	79	163	10	29
90	861	11	157	5	11
93	792	39	102	70	8
88	476	55	100	1	4
3	415	57	94	98	3
6	393	37	86	78	2
91	335	20	84	8	1
100	297	45	79	75	1
89	263	92	47	76	1
28	259	42	37	99	1
87	211	97	36		
56	209	4	34		

Quadro 11 – Quantitativo de neurônios nos agrupamentos, após a classificação do distrito 15

4.8 Sugestão de atividades corretivas para aumento de arrecadação

A partir dos agrupamentos identificados, e da verificação de que grupos de imóveis estão potencialmente desatualizados, ações específicas de atualização, com ou sem procedimento fiscal, podem ser adotadas, envolvendo, especialmente, os auditores e assistentes técnicos do tesouro, com maior habilidade na manipulação de dados em formato eletrônico, e na operação de computadores.

Também pode haver um direcionamento específico, a fim de que sejam identificados muitos dos proprietários de imóveis atualmente desconhecidos, com trabalho específico para os imóveis cujas distorções causam maior perda de receita tributária para o município, ou seja, trabalhar a atualização dos nomes dos proprietários, preferencialmente iniciando pelos imóveis selecionados pela Rede Neural como fora de agrupamentos padrões.

Agrupamentos após a classificação

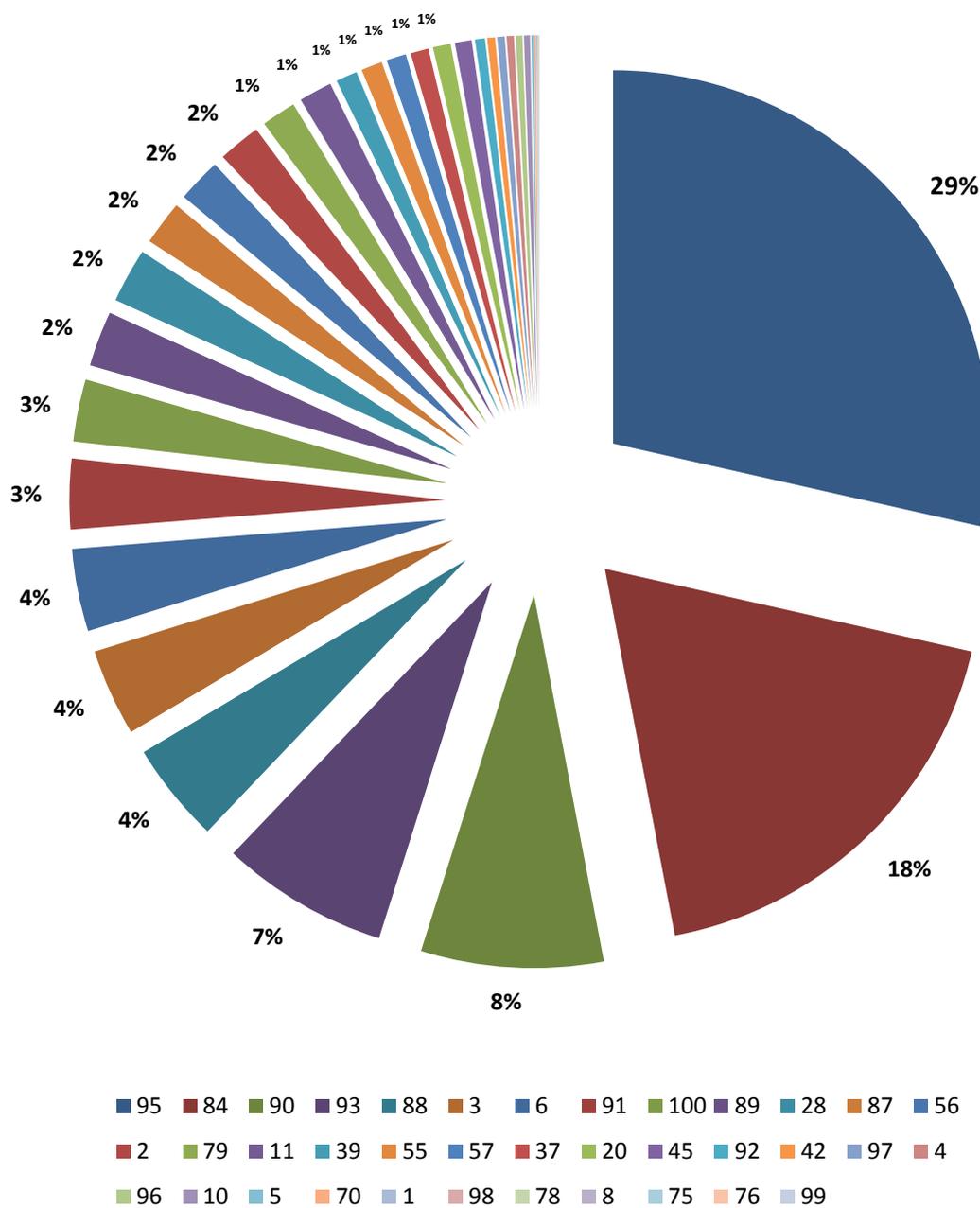


Figura 31 – Agrupamentos formados após a classificação do distrito 15

Considerando que a rede pode ser treinada de forma direcionada, ou seja, selecionando-se os neurônios que se deseja utilizar para a composição da Camada de Entrada, trabalhos específicos podem ser realizados para potencializar a utilização dos recursos humanos e materiais, a fim de obter maior produtividade nas tarefas desenvolvidas.

Como exemplo, é possível citar a atualização das datas de construção dos imóveis, pois tal variável é utilizada para o cálculo da depreciação aplicada aos imóveis edificados. Quando a data de nova construção não é corretamente informada ao Fisco, pode passar a ocorrer expressiva perda de receita tributária, uma vez que o fator de depreciação, em geral, deixa de ser atualizado. Sem o uso de metodologia específica, como as Redes Neurais Artificiais, o trabalho de atualização das datas de construção é priorizado a partir de fatores simples como a área edificada do imóvel, ou de uma composição entre dois ou mais fatores, como a área edificada e o uso específico (residencial ou não residencial). A utilização da metodologia objeto deste trabalho traz, assim, um estabelecimento mais eficaz de prioridades e, conseqüentemente, um maior rendimento das atividades da equipe de auditores e assistentes técnicos, desaguando em relevante acréscimo na arrecadação do tributo.

5 TRABALHOS CORRELATOS

5.1 Considerações iniciais

O objetivo deste capítulo é apresentar trabalhos versando sobre a aplicação de técnicas de clusterização em temas relativos a imóveis. Em busca realizada na internet em 15 de maio de 2010, foram encontrados poucos trabalhos específicos englobando os dois temas, Redes de Kohonen e avaliação imobiliária em massa para fins de IPTU. No entanto, de forma genérica, as Redes Neurais Artificiais foram objeto de vários trabalhos relacionados à avaliação imobiliária. Alguns deles são apresentados adiante.

5.2 Revisão bibliográfica

Dois artigos versando sobre a aplicação prática das RNA são ampla e detalhadamente comentados nos itens que se seguem. Também é feita referência a uma tese de doutorado e a duas dissertações de mestrado tratando de assunto correlato.

5.2.1 Trabalho 1: “Análise da utilização de inteligência artificial na determinação de plantas de valores”

Em artigo intitulado “Análise da utilização de inteligência artificial na determinação de plantas de valores”, González e Formoso (2000) abordam o tema de avaliação imobiliária considerando que a subjetividade das técnicas utilizadas na avaliação imobiliária deixam de considerar de forma eficiente os efeitos espaciais em decorrência da complexidade do mercado imobiliário. Frisam que é importante buscar alternativas de cálculo em busca da justiça tributária, e apresentam duas técnicas de inteligência artificial, redes neurais e raciocínio baseado em casos, estudando suas possíveis aplicações na definição de uma Planta Genérica de Valores Imobiliários (PGVI).

Apesar da sólida base estatística, a Análise de Regressão Múltipla (ARM) empregada há vários anos na avaliação de imóveis não tem oferecido soluções

adequadas para os problemas apontados, gerando a necessidade de métodos mais avançados ligados à Inteligência Artificial (IA): as Redes Neurais Artificiais (RNA) e os Raciocínios Baseados em Casos (RBC).

Após estabelecer relação com os neurônios cerebrais, os autores comentam sobre a utilidade das RNA, cujo emprego vai da identificação de bons e maus pagadores até a avaliação de imóveis, e detalham o mecanismo de funcionamento da rede, destacando que “as RNA podem ser entendidas como mecanismos de reconhecimento de padrões, com habilidade de autoaprendizagem” (GONZÁLEZ; FORMOSO, 2000, p.2). São apresentados formatos distintos de redes: a Perceptron Multicamadas com treinamento supervisionado possui formato com três camadas. Já os Mapas Auto-organizáveis de Kohonen, cujo treinamento é não-supervisionado, possui duas camadas. É comentado também o algoritmo de treinamento *backpropagation*, quando a rede é iniciada com configuração complexa até obter a necessária simplificação.

Menciona-se, no artigo em questão, que existe semelhança entre alguns modelos de redes neurais e a análise de regressão padrão através de modelos generalizados. É lembrado, ainda, que pode ocorrer o fenômeno conhecido como **supertreinamento**, quando a rede ultrapassa os limites ótimos de aprendizado e memoriza os dados, em cujo caso, o treinamento se aplicará apenas ao conjunto de dados a ele submetido.

Os autores pontuam ainda que, em alguns países da Europa, as RNA não podem ser utilizadas, por exemplo, para fins de análise de crédito, pois são necessárias regras claras para a negativa de crédito, e o mecanismo de funcionamento de tais redes equivale a uma função ponderada dos vetores, sem lógica clara que demonstre os motivos pelos quais se chegou a um determinado resultado.

No mesmo trabalho, os autores expõem o Raciocínio Baseado em Casos, técnica que pretende resolver problemas através do conhecimento e análise de observações anteriores de casos análogos. O RBC se baseia nas premissas de que “problemas similares têm soluções similares, e a reutilização de soluções de casos anteriores é mais adequada do que uma solução baseada na generalização de regras” (GONZÁLEZ; FORMOSO, 2000, p.6). São apontadas, como vantagens da técnica, o fato de não depender de modelo explícito para a solução do problema, e de ser de fácil atualização.

A conclusão dos autores é apresentada no sentido de que tanto as RNA quanto os RBC são bastante úteis na análise dos valores dos imóveis, desde que com os devidos ajustes. No caso das Redes Neurais Artificiais, as maiores dificuldades encontradas foram a baixa velocidade de processamento dos computadores disponíveis na época, bem como a necessidade de um especialista no assunto para a definição dos critérios de parada. Citando Watson (1997), ressaltam que “as RNA são melhores para áreas onde os dados não podem ser representados simbolicamente”. Já o Raciocínio Baseado em Casos se mostrou eficiente nos casos de dados simbólicos complexos e estruturados, mas deficiente quanto à recuperação e à adaptação de casos.

5.2.2 Trabalho 2: “Da teoria à prática - introduzindo técnicas emergentes na avaliação de imóveis em massa”

Em artigo intitulado “Da teoria à prática - introduzindo técnicas emergentes na avaliação de imóveis em massa”, Brondino e Silva (2005) escrevendo na Revista Minerva discorrem sobre o tema da avaliação imobiliária, inicialmente traçando, como objetivos do trabalho, a apresentação de alternativas para a avaliação de imóveis em massa.

Explicam os autores que a avaliação imobiliária, na atualidade, deixa de ser assunto apenas de engenheiros. É multidisciplinar. No escopo da avaliação, estão desde aspectos de engenharia até aspectos econômicos e sociais, como a renda dos vizinhos e as facilidades de transporte. Mas inúmeras variáveis precisam ser analisadas, inclusive as de natureza espacial ainda pouco utilizadas no Brasil.

É explicado que a eficiência do método de avaliação é tão importante quanto as variáveis a serem utilizadas. A dificuldade de leitura de modelos estatísticos complexos, com grande número de variáveis e difícil implementação, restringe o uso a profissionais altamente qualificados. É o caso da Regressão Múltipla utilizada largamente na avaliação imobiliária em massa e que, embora eficiente, pode apresentar problemas de descontinuidade e não-linearidade capazes de afetar a eficácia de todo o processo.

Desta forma, os autores, citando Trippi e Turban (1992) e Luger e Stubblefield (1989), apresentam, como alternativa bastante atraente, as Redes Neurais Artificiais, cujas principais vantagens apontadas são: processamento em

tempo real, tolerância a erros nas entradas e desnecessidade de imposição de algumas condições existentes nos métodos estatísticos convencionais. Destacam, ainda, que, “apesar das vantagens mencionadas, há relativamente poucos exemplos na literatura que fizeram uso das Redes Neurais, para avaliar propriedades”, citando Guedes (1995), Almond, Jenkins e Gronow (1997), Brondino e Silva (1998).

A abordagem proposta foi utilizada na cidade de São Carlos, município a 230 Km de São Paulo capital, onde existem muitos vazios urbanos, conforme Silva (1997) citado no artigo. Foram combinados recursos de Sistemas de Informações Geográficas (SIG) com Redes Neurais Artificiais, no entanto apenas estas últimas foram comentadas pelos autores que justificaram a omissão dos comentários sobre os SIG em virtude de julgarem já ser de ampla aplicação nos meios técnicos e acadêmicos.

No texto, é descrita a composição de um neurônio, e é comparado o funcionamento do cérebro em relação ao dos computadores, destacando as vantagens e desvantagens de cada um, e explicando como funciona o paralelismo cerebral cujo comportamento os computadores buscam simular através das redes neurais. É explicado que a topologia da rede diz respeito à forma como os neurônios se comunicam, o que afeta o desempenho da rede.

Algumas das redes neurais permitem que o neurônio envie sinais de volta para ele mesmo, frisam os autores, citando Tubb (1993). É o caso do tipo de rede mais comum, a *Multilayer Perceptron* (MLP), composta por uma camada de entrada, uma ou mais camadas intermediárias e uma camada de saída. Após a descrição do funcionamento das redes MLP são apresentadas as etapas da metodologia de avaliação: primeiramente são escolhidas as variáveis relacionadas aos imóveis. Em seguida executa-se a amostragem dos terrenos e suas digitalizações no mapa da cidade a partir de um SIG. Na terceira etapa, define-se o modelo de Rede Neural a utilizar. Com a melhor configuração da rede definida, identifica-se a influência de cada variável no resultado final, descartando-se variáveis irrelevantes e fazendo novo treinamento da rede.

Fazendo a sobreposição de dois mapas, o de bairro e um segundo obtido a partir de informações censitárias do Instituto Brasileiro de Geografia e Estatística (IBGE), obteve-se a interseção renda-bairro após a segmentação dos níveis de renda em três classes distintas. A medida de acessibilidade, variável correspondente à distância do imóvel até o centro da cidade, foi também definida.

Reportam os autores que com o *software* Sistema de Informações Geográficas para Transportes TransCAD, citado por Caliper (1996), a localização dos imóveis foi inserida no mapa da cidade. Em seguida, utilizando o *software Neural Planner* versão 4.2, a simulação da rede usando o modelo MLP foi processada, tendo sido precedida dos ajustes necessários para a adaptação das variáveis não numéricas, como, por exemplo, a existência de muro, calçada, rede de esgoto e outras, que foram convertidas para “0”, representando a resposta “não”, ou para “1”, representando “sim”. Diversas simulações foram realizadas com configurações distintas da rede para cada simulação. Aquela que apresentou o menor erro foi a escolhida.

Citando Nath, Rajagolapalan e Ryker (1997), Brondino e Silva entendem que a relevância de cada variável no modelo pode ser analisada segundo o método proposto por Garson (1991). Identificadas as variáveis mais relevantes, apenas estas são utilizadas para o treinamento definitivo da rede. No caso proposto pelos autores, são desprezadas as variáveis cuja representação seja de no máximo 5%, conforme cálculo de representatividade específico descrito no artigo. Revelaram-se não representativas as variáveis topografia, situação, frente do terreno e asfalto.

Os autores concluem que a abordagem proposta mostrou-se capaz de substituir os modelos estatísticos convencionais, indicando que as Redes Neurais são uma boa alternativa para a avaliação dos imóveis, permitindo ainda a análise de relevância de cada variável na composição do valor imobiliário. Além do mais, as RNA permitem grande facilidade de operação por um usuário comum de computadores, exigindo apenas o conhecimento dos procedimentos e da parametrização necessária para a utilização da Rede Neural Artificial.

5.2.3 Outros trabalhos

A seguir são listados outros trabalhos relacionados ao tema Redes Neurais Artificiais e que tratam da avaliação imobiliária:

- a) Tese de doutorado apresentada, em 1999, por Nair Cristina Margarida Brondino à Escola de Engenharia de São Carlos, da Universidade de São Paulo, como requisito para a obtenção do título de Doutor em Engenharia Civil – Transportes sob o título “Estudo da influência da acessibilidade no valor de lotes urbanos através do uso de Redes Neurais” (BRONDINO, 1999).

- b) Dissertação de mestrado apresentada, em 2007, por Roberto Ruano Dalaqua ao programa de pós-graduação da Faculdade de Ciências e Tecnologia da Universidade Estadual Paulista – UNESP, campus de Presidente Prudente-SP, como requisito para a obtenção do título de Mestre em Ciências Cartográficas sob o título “Aplicação de métodos combinados de avaliação imobiliária na elaboração da Planta de Valores Genéricos” (DALAQUA, 2007).
- c) Dissertação de mestrado apresentada, em 2006, por Antônio Pelli Neto ao programa de pós-graduação da Universidade Federal de Minas Gerais, como requisito para a obtenção do título de Mestre em Engenharia Elétrica sob o título “Redes neurais artificiais aplicadas às avaliações em massa - estudo de caso para a cidade de Belo Horizonte / MG” (PELLI NETO, 2006).

5.3 Considerações finais

Alguns trabalhos relacionados ao tema Redes Neurais Artificiais foram expostos ao longo deste capítulo. Uns abordando a avaliação imobiliária de forma geral, outros mencionando especificamente aspectos ligados ao IPTU, mas todos, de uma forma ou de outra, abordando a aplicação da tecnologia objeto desta dissertação, que já tem sido vista por muitos autores e estudiosos do tema como uma opção de substituição da antiga metodologia estatística de valoração da malha urbana.

Percebe-se que, desde artigos até teses de doutorado, podem ser encontrados na literatura com larga análise e prova de utilidade da aplicação ao caso concreto. Observa-se também que, pouco a pouco, os autores sugerem a substituição dos complexos modelos estatísticos pela aplicação profissional efetiva das RNA, independente da metodologia que adotem, convencendo-se de que é uma opção mais moderna, eficiente e de mais fácil utilização que as antigas metodologias.

6 CONCLUSÕES

As Redes Neurais Artificiais são metodologia acessível e de largo espectro de aplicação na sociedade. Prestam-se à identificação de grupos semelhantes, não facilmente identificáveis pela lógica do tipo “se → então” com que usualmente se está acostumado a trabalhar, tanto no universo avançado das planilhas eletrônicas, quanto nos programas de computador mais complexos.

O processamento em paralelo, copiado pelas RNA do mecanismo verificado no córtex cerebral, permite simular operações de elevado grau de complexidade, através das quais, o computador pode aprender com o próprio ambiente, com os fatos acontecidos e com as características dos objetos apresentados à rede.

As áreas em que as Redes Neurais Artificiais podem ser utilizadas vão, desde aplicações industriais ou comerciais, a públicas ou privadas, civis ou militares. Demonstrou-se, neste trabalho, um estudo de caso que permite à Administração Tributária identificar distorções causadoras de perdas de receita para o ente estatal, com aplicação prática.

Sendo a metodologia, objeto de nosso estudo, aplicável à classificação e agrupamento de quaisquer espécies de variáveis, contínuas ou discretas, representativas de elementos constantes nos mais diversos bancos de dados, abre-se um vasto leque de aplicações para sua utilização. Certamente já há bastante tempo as RNA vêm sendo utilizadas para a identificação de padrões, mas aqui pretendemos exaltar, que a simplicidade e variedade de seu uso podem trazê-la para a solução de muitos problemas cotidianos do serviço público civil ou militar, tanto quanto de empresas, prestadores de serviço, ou das indústrias em geral.

Mais uma vez lembramos que o aprendizado da Rede acontece de forma diversa da lógica tradicional, daí o valor agregado que se pode conseguir nas mais diversas áreas para as quais existam dados que se deseje agrupar com a eficiente lógica de funcionamento do cérebro humano como, por exemplo, no(a):

- a) adequação de equipe multidisciplinar de profissionais;
- b) alocação de veículos para o transporte público;
- c) análise de risco de aplicações financeiras;
- d) aprovação de crédito bancário;

- e) controle da saúde pública por meio do cruzamento de dados sobre as áreas de risco, o perfil de pessoas infectadas e os investimentos públicos correspondentes;
- f) definição de grupos, homogêneos ou não, para treinamentos civis ou militares;
- g) eleição de questões para concursos públicos;
- h) identificação de áreas para investimentos (construções, prospecções de petróleo, instalação de equipamentos públicos, telefonia, energia, centros comerciais e outras);
- i) identificação de perfis de uso para a indústria (automobilística, moveleira, de calçados e tantas outras), a fim de auxiliar no planejamento da produção;
- j) identificação do perfil de criminosos;
- k) obtenção de indícios de sonegação fiscal;
- l) planejamento de vendas no varejo ou no atacado;
- m) previsão de falhas em máquinas industriais;
- n) seleção de pessoal;

A flexibilidade de parametrização da Rede Neural de Kohonen torna o treinamento extrema e facilmente ajustável a cada necessidade, tal como implementado no *software* Analisador do Solo Urbano desenvolvido pelo mestrando, considerando os dados fornecidos pela Secretaria de Finanças do Município de Fortaleza. A operação do *software* é elementar para usuários com algum conhecimento básico sobre as RNA e sobre microcomputadores. As possibilidades de utilização da tecnologia apresentada são inúmeras, e, estabelecer a cultura de seu uso, para fins tributários, ou em qualquer outro meio onde se aplique, tende a facilitar a implantação e utilização, com respostas rápidas e eficazes.

A utilização das RNA, embasando o direcionamento de atualizações cadastrais em bancos de dados imobiliários, otimiza, sobremaneira, a eficácia das ações fiscais. O modelo de Kohonen, em trabalhos futuros, pode ser substituído por outro, que se julgue mais adequado para a classificação, tal como os algoritmos genéticos, por exemplo, sendo viável, inclusive, o desenvolvimento de modelo próprio que, na sua essência, incorpore elementos que confirmam especial habilidade tributária ao mecanismo de treinamento, ajudando a alavancar a arrecadação do IPTU no município de Fortaleza, que alcançou R\$ 114.115.675,26 em 2009.

Uma vez que o Analisador foi desenvolvido com viés acadêmico, pretende-se implementar, oportunamente, uma versão profissional que contemple não apenas outros tributos, mas que seja adaptável a realidades de outros municípios por meio de parametrizações e, tanto quanto possível, a depender da relação custo/benefício, que abra espaço também para seu uso pela iniciativa privada. Vislumbramos, assim, um largo espectro de aplicações para as Redes Neurais Artificiais que, uma vez utilizadas conjuntamente com outras metodologias de seleção de dados, podem trazer expressivo ganho no apoio a decisões.

No município de Fortaleza, através de sua Administração Fiscal, especificamente no tocante à identificação de distorções tributárias para fins de priorização das ações de fiscalização e atualização cadastral de imóveis, a utilização das RNA, em conjunto com métodos tradicionais de seleção e análise de dados imobiliários, tem-se mostrado eficiente, prática e com ótimas perspectivas de acréscimo de arrecadação. Evidentemente que o Operador da Rede, aquele que a parametriza com a habilidade técnica de conhecedor das regras de negócio que deve ser, precisa ser cuidadoso o suficiente para não inviabilizar o aprendizado, tornando os resultados pouco expressivos. Por exemplo, se o Operador define a quantidade de classes para a classificação dos grupos, com um quantitativo inferior ao número de valores passíveis de serem assumidos por uma determinada variável, ele pode estar condenando a Rede ao insucesso. Assim, se no município de Fortaleza existem 13 classificações arquitetônicas (casa, apartamento, galpão, subsolo, loja e sala, dentre outras), ao definir uma quantidade de 4 classes para as possibilidades de entrada, fatalmente não haverá espaço para uma minuciosa identificação e diferenciação entre uma casa e um apartamento, ou entre um galpão e um subsolo, condenando o agrupamento final a uma indesejável miscigenação. Daí porque a eficiência do aprendizado, está intimamente ligada ao conhecimento do Operador da Rede para operá-la, sendo indispensável a devida atenção no estabelecimento de seus parâmetros. Bem configurada, a Rede surpreende.

BIBLIOGRAFIA

ALEXANDRE, R. **Direito tributário esquematizado**. 2. ed. São Paulo: Método, 2008.

ALMOND, N.; JENKINS, D.; GRONOW, S. Development of a prototype residential valuation system. In: **COMPUTERS IN URBAN PLANNING AND URBAN MANAGEMENT**, 5., 1997, Mumbai. **Proceedings...** Mumbai, 1997. v. 1, p. 204-215.

ANDRADE, H. da S. **Um processo de mineração de dados aplicado ao combate à sonegação fiscal do ICMS**. 2009. 161f. Dissertação (Mestrado Profissional em Computação) - UECE, Fortaleza.

BAPTISTELA, M. **O uso de redes neurais e regressão linear múltipla na engenharia de avaliações**: determinação dos valores venais dos imóveis urbanos. 2005. 123f. Dissertação (Mestrado em Métodos Numéricos em Engenharia)– Universidade Federal do Paraná, Curitiba, 2005.

BEALE, R.; JACKSON, T. **Neural computing**: an introduction. New York: Taylor & Francis, 1990.

BRAGA, A. de P.; CARVALHO, A. C. P. L. F.; LUDEMIR, T. B. **Redes neurais artificiais** – teoria e aplicações. Rio de Janeiro: LTC, 2000.

BRONDINO, N. C. M. **Estudo da influência da acessibilidade no valor de lotes urbanos através do uso de redes neurais**. 1999. 158f. Tese (Doutorado em Engenharia Civil – Transportes)–Escola de Engenharia de São Carlos, Universidade de São Paulo – USP, São Carlos.

BRONDINO, N. C. M; SILVA, A. N. R. Da teoria à prática: introduzindo técnicas emergentes na avaliação de imóveis em massa. **Revista Minerva**, São Carlos - SP, v. 2, n. 2, p. 107-117, 2005.

_____. A comparison of land valuation methods supported by GIS. In: **INTERNATIONAL CONFERENCE ON DESIGN AND DECISION SUPPORT SYSTEMS IN ARCHITETURE AND URBAN PLANNING**, 1998, Maastricht. **Proceedings...** Maastricht, 1998. 1 CDROM.

CALDEIRAS, A. M. et al. **Inteligência computacional aplicada à administração, economia e engenharia em Matlab**. São Paulo: Thomson Learning, 2007.

CALIPER. **TransCAD** – Transportation GIS Software – manual supplement version 3.0. Newton, MA: Caliper Corporation, 1996.

CAMPELLO, R. E; MACULAN, N. **Algoritmos e heurísticas**: desenvolvimento e avaliação de performance. Niterói: Universidade Federal Fluminense, 1994.

CARVALHO, L. A. V. **Datamining** – a mineração de dados no marketing, medicina, economia, engenharia e administração. Rio de Janeiro: Ciência Moderna, 2005.

CURTY, M. G.; CRUZ, A. da C. **Guia para apresentação de trabalhos acadêmicos, dissertações e teses**. Maringá: Dental Press, 2001.

DALAUQUA, Roberto R. **Aplicação de métodos combinados de avaliação imobiliária na elaboração da planta de valores genéricos**. 2007. 128f. Dissertação (Mestrado em Ciências Cartográficas)–Estadual Paulista, Presidente Prudente, 2007.

ECO, Umberto. **Como se faz uma tese**. 20. Ed. São Paulo: Perspectiva, 2005.

ENSSLIN, L.; MONTIBELLER NETO, G.; NORONHA, S. M. **Apoio à decisão: metodologia para estruturação de problemas e avaliação multicritério de alternativas**. Florianópolis: Insular, 2001.

FERNANDES, C. E. **IPTU – Texto e contexto**. São Paulo: Quartier Latin, 2005.

FONSECA, J. S. da; MARTINS, G. de A. **Curso de estatística**. 6. ed. São Paulo: Atlas, 1996.

GARSON, D. G. Interpreting neural-network connection weights. **AI Expert**, p. 47-51, abr. 1991.

GÓES, H. **Elementos básicos de estatística**. Fortaleza: Universidade Estadual do Ceará, 1998.

GOMES, F. J.; MESQUITA, C. K. **Legislação tributária do município de Fortaleza (consolidação) & processo administrativo fiscal**. Fortaleza: Garin, 2003.

GONZÁLEZ, M. A. S.; FORMOSO, C. T. **Análise da utilização de inteligência artificial na determinação de plantas de valores**. Rio de Janeiro, 2001. Disponível em <http://www.pellisistemas.com.br/downloads/Inteligencia_Artificial_Planta_Valores.pdf>. Acesso em: 15 maio 2010.

GUEDES, J. C. O. Emprego de inteligência artificial na avaliação de bens. In: CONGRESSO BRASILEIRO DE ENGENHARIA DE AVALIAÇÕES E PERÍCIAS, 8., 1995, Florianópolis. **Anais...** Florianópolis, 1995. p. 368-374.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001.

IMAGUIRE, G.; BARROSO, C. A. C. **Lógica – os jogos da razão**. Fortaleza: UFC, 2006.

KOHONEN, T. The self-organizing map. **Proceedings of the IEEE**, v. 78, n. 9, 1990.

LACHTERMACHER, G. **Pesquisa operacional na tomada de decisões: modelagem em Excel**. 2. ed. Rio de Janeiro: Campus, 2004.

LEITHOLD, L. **O cálculo com geometria analítica**. São Paulo: Harper & Row, 1977. v. 1.

LIMA NETO, V. C.; ARAGÃO, J. J. G.; YAMASHITA, Y. Desenvolvimento de uma metodologia para estimar a mais-valia imobiliária decorrente de intervenções em infra-estrutura de transporte público metroviário. In: CONGRESSO BRASILEIRO DE TRANSPORTE E TRÂNSITO, 16., 2007, Maceió. **Anais...** Maceió, 2007, 10f.

LUGER, G. F.; STUBBLEFIELD, W. A. **Artificial intelligence and the design of expert systems**. The Benjamin/Cummings Publishing Company, 1989. p. 581-584.

MACHADO, H. de B. **Crimes contra a ordem tributária**. São Paulo: Atlas, 2008.

MACHADO, H. de B. **Curso de direito tributário**. 26. ed. São Paulo: Malheiros, 2005.

MELLO, C. A. B. **Curso de direito administrativo**. 26. ed. São Paulo: Malheiros, 2009.

MEYER, P. L. **Probabilidade – aplicações à estatística**. 2. ed. Rio de Janeiro: LTC, 1983.

MORAES, A. de. **Direito constitucional**. 8. ed. São Paulo: Atlas, 2000.

NATH, R.; RAJAGOLAPALAN, B.; RYKER, R. Determining the saliency of input variables in Neural Network classifiers. **Computers & Operational Research**, v. 24, n. 8, p. 767-773, 1997.

OJJA, E.; KASKI, S. **Kohonen Maps**. Amsterdam: Elsevier, 1999.

OLIVEIRA JR., H. A. e; CALDEIRA, A. M.; MACHADO, M. A. S SOUZA, R. C.; TANSCHIT, R. **Inteligência computacional aplicada à administração, economia e engenharia em Matlab®**. São Paulo: Thomson Learning, 2007.

PELLI NETO, A. **Redes neurais artificiais aplicadas às avaliações em massa: estudo de caso para a cidade de Belo Horizonte / MG**. 2006, 111p. Dissertação (Mestrado em Engenharia Elétrica) – Universidade Federal de Minas Gerais – UFMG, Belo Horizonte.

PINTO, A. L. T.; WINDT, M. C. V. S.; CÉSPEDES, L. (Org.). **Código tributário nacional e Constituição Federal**. 37. ed. São Paulo: Saraiva, 2008.

_____. **Código penal**. 19. ed. São Paulo: Saraiva, 2004.

PORTELA, P. de O. **Apresentação de trabalhos acadêmicos de acordo com as normas de documentação da ABNT: informações básicas**. Uberaba: UNIUBE, 2005.

PRIDDY, K. L.; KELLER, P. E. **Artificial neural networks: an introduction**. Washington: SPIE, 2005.

RAO, M. A.; SRINIVAS, J. **Neural networks**. United Kingdom: Alpha Science International, 2003.

SEVERINO, Antônio J. **Metodologia do trabalho científico**. 23. ed. São Paulo: Cortez, 2009.

SILVA, A. N. R. Os custos do não-planejamento em São Carlos. In: SEMINÁRIO PLANO DIRETOR DE SÃO CARLOS – PROJETO CIDADE URGENTE, 1997, São Carlos. **Anais...** São Carlos: Universidade Federal de São Carlos, 1997. p. 51-54.

SILVA, E. M.; SILVA, E. M.; GONÇALVES, V.; MUROLO, A. C. **Estatística para os cursos de: economia, administração, ciências contábeis**. 3. ed. São Paulo: Atlas, 1999. v. 1.

SILVA, J. S. V. **Avaliação de imóveis urbanos para fins de gestão tributária municipal**: abordagem multicritério com vista à elaboração de uma planta genérica de Valores. 2006. Tese (Doutorado em Engenharia e Gestão Industrial)– Universidade Técnica de Lisboa, Lisboa, PT, 2006.

SOARES, A. C. **Pós-processamento de regras de associação e CRISP-DM na descoberta de conhecimento em dados de ocorrências policiais**. 2009. 138f. Dissertação (Mestrado integrado profissional em Computação Aplicada)– Universidade Estadual do Ceará, Fortaleza 2009.

SPÄTH, H. **Cluster Analysis Algorithms for data reduction and classification objects**. England: Ellis Horwood, 1980.

TRIPPI, R. R.; TURBAN, E. **Neural Networks in finance and investing**. [S.l.]: Probus Publishing Company, 1992. p. 3-72.

TUBB, N. R. A development path to success in neural computing. **Expert Systems Applications**, v. 9, n. 5, p. 5-9, 1993.

VIANA, G. V. R. **Meta-heurísticas e programação paralela em otimização combinatória**. Fortaleza: UFC, 1998.

WAGNER, H. M. **Pesquisa Operacional**. 2. ed. Rio de Janeiro: Prentice-Hall, 1986.

WATSON, I. **Applying case-base reasoning**: techniques for enterprise systems. San Francisco (USA): Morgan Kaufmann, 1997.

ZUBEN, F. J. Von; ATTUX, R. R. F. **Rede neural de Kohonen e aprendizado não-supervisionado**. São Paulo: Unicamp, [200-?].

GLOSSÁRIO

Ajuste de pesos

É o procedimento, através do qual, o neurônio vencedor e seus vizinhos têm corrigidas as forças de suas conexões, a fim de refletir uma maior afinidade com as entradas.

Através do ajuste dos pesos, a Rede Neural adquire conhecimento, pois é, através dele, que ocorre o aprendizado que qualifica a rede a classificar adequadamente as entradas de acordo com as saídas.

Alíquota

Valor percentual que, multiplicado pela base de cálculo, resulta no valor do imposto a pagar.

Aprendizado

Decorre do *treinamento* e confunde-se com este, dotando a rede do conhecimento necessário e suficiente, para classificar as entradas de acordo com as saídas.

Base de cálculo

Valor que serve de fundamento para o cálculo do imposto. No caso de imóveis, a base de cálculo do IPTU é o seu valor venal.

Camada de competição

É o nome dado à Camada de Saída no treinamento não-supervisionado, pois, neste tipo de treinamento os neurônios da camada competem entre si, para que seja identificado qual deles tem maior semelhança com a entrada apresentada.

Camada de entrada

É a camada formada pelo conjunto de dados de entrada, ou seja, os dados que são apresentados à Rede Neural para processamento. Normalmente compõem a camada de entrada os dados selecionados a partir do banco de dados através do pré-processamento.

Camada de Saída

Mesmo que Camada de Competição.

Cartografia

É a coordenada que indica em qual distrito, quadra, lote e complemento do lote, um determinado imóvel está localizado.

Por exemplo, se um imóvel está localizado no distrito 10, quadra 65, lote 9 e complemento do lote 88, então sua cartografia será 10.65.9.88. O complemento do lote indica que o imóvel indicado na cartografia é o n-ésimo imóvel naquele lote. No exemplo anteriormente apresentado, o imóvel em questão é o 88º do lote, correspondendo, provavelmente, a um imóvel em um bloco de apartamentos ou salas comerciais.

Ciclo

Cada ciclo corresponde a uma apresentação de todas as entradas para treinamento pela Rede Neural, ou seja, de todo o conjunto de dados.

Após o treinamento com o último elemento da Camada de Entrada, um novo ciclo é iniciado, tomando-se novamente todos os elementos, um a um, e apresentando-os à rede para seu aprimoramento. Ciclos são também chamados de iterações.

Clusterizar

Significa agrupar elementos similares a partir da comparação de um conjunto de características fornecidas juntamente com cada entrada, após o devido treinamento da rede.

A clusterização deve ser parametrizada pelo Operador da Rede, que deve definir, especialmente, a quantidade máxima de grupos a serem formados na Camada de Saída e os critérios de parada que causarão o final do treinamento.

Complemento do lote

É a parte da cartografia que define a existência de subunidades dentro de um mesmo lote de terra.

Ocorre quando, por exemplo, são edificadas vários apartamentos em um determinado condomínio ou várias lojas em um determinado prédio comercial. Seja, por exemplo, uma cartografia indicada por 10.65.9.88, tem-se que o número mais à direita indica o complemento do lote. No exemplo dado, a cartografia apresentada especifica a 88ª unidade dentro do mesmo lote.

Convergência

Ocorre, quando o treinamento da Rede Neural não está levando mais a qualquer aprendizado.

Diz-se que a rede, ao alcançar a convergência, está estável. Ou seja, ao serem apresentadas novas entradas, não mais acontecem alterações nos pesos das conexões entre a Camada de Entrada e a Camada de Saída.

Córtex cerebral

É o núcleo do cérebro, da inteligência.

Possui altíssima densidade de neurônios e responde pela capacidade de aprendizado e adaptação. Localiza-se nos hemisférios cerebrais e representa a camada mais complexa de seu funcionamento.

Critério de parada

É a condição sob a qual o treinamento da rede é interrompido e a rede é considerada treinada ficando pronta, para ser submetida à efetiva classificação das entradas.

Para evitar que o treinamento da Rede Neural seja excessivamente demorado, o Operador da Rede estabelece critérios que, alcançados, darão a rede por treinada. Alguns critérios de parada podem ser, por exemplo: tempo de processamento, quantidade de ciclos processados e falta de alteração nos pesos dos neurônios após N entradas. Outros critérios de parada podem ser implantados pelo Operador da Rede, como conveniente.

Dendritos

São filamentos irregulares, parte da estrutura celular, através dos quais são transportadas informações entre neurônios, para conduzi-las ao corpo da célula.

Distrito

Maior subdivisão do município. É parte da cartografia e o polígono que o limita é usualmente estabelecido em lei.

Um distrito é formado por várias quadras ou quarteirões. Estas quadras nem sempre são polígonos regulares nem tampouco de quatro lados. Em virtude da dinamicidade na criação de novos bairros, estes findam por não coincidir com os limites dos distritos. Assim, é comum encontrar um bairro contendo partes de distritos distintos, bem como distritos contendo partes de bairros distintos.

Se uma cartografia é indicada por 10.65.20.88, tem-se que o primeiro número mais à esquerda, ou seja, 10, é o número do distrito. No exemplo dado, a cartografia apresentada especifica o 10º distrito em que o município foi subdividido.

Erros mínimos quadrados

Nome pelo qual também é conhecida a regra Delta de aprendizado supervisionado, quando tanto a entrada quanto a saída desejada são conhecidas e o treinamento da Rede Neural ocorre no sentido de encontrar o

ajuste de pesos ideal, a fim de que as entradas apresentadas encontrem correspondência na Camada de Saída.

Estabilidade

Estado em que se acha a rede, quando o treinamento não está mais levando a qualquer aprendizado, ou seja, quando alcançou a convergência.

Uma rede estável não mais vê alterados os pesos das conexões entre a Camada de Entrada e a Camada de Saída, quando são apresentadas novas entradas.

Inscrição imobiliária

É um número pelo qual se faz referência ao imóvel.

Um condomínio vertical de uso comercial, para o qual existam 70 salas, possui 70 inscrições no Cadastro Imobiliário. No aspecto tributário e também no presente estudo, uma inscrição é tratada como um imóvel independente, possuindo características únicas a ela relacionadas. As inscrições têm o formato $NNNNN-D$ onde N é um algarismo de 0 a 9, e D é um dígito verificador também representado com um algarismo de 0 a 9.

Iteração

Mesmo que ciclo.

Lote

É a parte da cartografia que indica as parcelas de terra nas quais foram subdivididas as quadras. Um lote é formado de uma ou mais unidades. Em caso de subdivisão do lote, cada subdivisão recebe um número chamado de complemento do lote.

Um lote de terra não edificado é cadastrado como terreno nos cadastros imobiliários. Se edificado, o lote pode conter uma casa, um galpão ou um bloco de apartamentos, por exemplo. Se uma cartografia é indicada por 10.65.20.88, tem-se que o segundo número mais à direita, ou seja, 20, é o número do lote. No exemplo dado, a cartografia apresentada especifica o 20º lote em que a quadra foi subdividida.

Mapas auto-organizáveis de Kohonen

Tipo de Rede Neural Artificial (RNA) cujo objetivo é conseguir que valores apresentados, como entrada, sejam automaticamente organizados em grupos de saída pré-definidos.

Sendo apenas as entradas conhecidas, a auto-organização é conseguida a partir de um mecanismo de treinamento não supervisionado que calcula e

ajusta pesos sinápticos representativos da similaridade entre os dados apresentados e os grupos de saída disponíveis, imitando o funcionamento do córtex cerebral. Seu idealizador foi o finlandês Teuvo Kohonen, pesquisador da Universidade de Helsinque, que criou o mecanismo na década de 80. São também conhecidos como a Rede Neural de Kohonen.

Memória associativa

É a capacidade que possuem os seres humanos de recuperarem informações armazenadas no cérebro, a partir da associação com outras informações igualmente armazenadas.

Por exemplo, quando uma determinada pessoa deseja lembrar do ano em que comprou o seu primeiro automóvel, através da memória associativa, poder-se-ia lembrar que o automóvel foi comprado no ano em que a Seleção Brasileira de Futebol ganhou o tricampeonato no México e daí deduzir que o automóvel foi comprado em 1970, por associação direta entre o ano da compra do automóvel e o ano da Copa do Mundo em questão.

Mineração de dados

É uma metodologia de busca e extração de dados em grandes bancos de dados, através de técnicas de inteligência artificial que procuram estabelecer relacionamentos entre as diversas variáveis do banco de dados.

Neurônio vencedor

É o que demonstra possuir maior similaridade com a entrada apresentada, dentre todos os neurônios da Camada de Saída.

Quando uma entrada – que pode ser composta por um ou mais neurônios dispostos na Camada de Entrada – é apresentada à rede, verifica-se a força das conexões entre a entrada e cada neurônio da Camada de Saída. Na comparação, é dito vencedor o neurônio da Camada de Saída com maior similaridade em relação à entrada apresentada.

Neurotransmissores

São elementos químicos liberados pela sinapse, quando os pulsos elétricos produzidos pelo axônio (canal de saída do neurônio) alcançam níveis suficientemente elevados.

Operador da rede

Aquele indivíduo ou usuário responsável pela parametrização da Rede Neural, antes que seja iniciado o treinamento.

É quem define, por exemplo, quantos neurônios irão compor a Camada de Saída, quais os critérios de parada, qual a topologia da vizinhança afetada pelo

neurônio vencedor, e qual o raio em que os neurônios serão considerados vizinhos.

Peso

Elemento indicador de similaridade entre as entradas e os neurônios da Camada de Saída, antes do treinamento, o peso é inicializado com valores baixos, sendo ajustado ao longo de todo o aprendizado.

Os pesos são os efetivos responsáveis pelo aprendizado da Rede Neural, permitindo que após o treinamento da rede, uma entrada adicional seja adequadamente classificada entre os neurônios (grupos) da Camada de Saída.

Planta genérica de valores imobiliários (PGVI)

É o documento através do qual os municípios oficializam os valores dos terrenos em uma determinada quadra, para fins de tributação.

É comum que a PGVI traga para uma mesma quadra, valores específicos para cada uma de suas faces, visto que estas podem apresentar valores distintos em virtude da maior ou menor proximidade de equipamentos urbanos ou elementos valorizantes ou desvalorizantes como hospitais, colégios, casas de espetáculos, delegacias de polícia, favelas, fluxo de veículos, postos de gasolina e outros.

Potencial de ativação

É a medida dos pulsos elétricos emitidos pelo axônio que, quando alcançada, retira o neurônio de seu estado de repouso.

Quando o nível do potencial de ativação é alcançado, a sinapse libera os neurotransmissores que fazem com que o axônio de um neurônio se comunique com o dendrito de outro, excitando ou inibindo o fluxo de informação na célula de destino.

Processamento distribuído

É a capacidade que possui a estrutura cerebral de dividir uma determinada tarefa entre vários neurônios.

Este tipo de processamento é também conhecido como processamento em paralelo, contrapondo-se à maneira seriada com que os computadores executam suas atividades, realizando uma tarefa apenas após o término da tarefa anterior. Através da distribuição de tarefas, ou do processamento, cria-se a possibilidade da estrutura se auto-ajustar, quando ocorrer a perda permanente de um neurônio, substituindo-o por outro, sem que haja prejuízos para a consecução da tarefa como um todo.

Quadra

É a parte da cartografia que indica as parcelas de terra nas quais foram subdivididos os distritos. Uma quadra é formada de um ou mais lotes.

Apesar do nome quadra sugerir quatro lados, nem sempre o polígono de uma quadra possui exatamente quatro faces, podendo possuir qualquer número a partir de três faces, quantidade mínima para que exista um polígono fechado, ainda que uma das faces seja inundável. Quando projetada de forma regular, a quadra normalmente possui 100m x 100m de área, com um total de 10.000 m² ou 1 ha, sendo popularmente conhecida como quarteirão.

Se uma cartografia é indicada por 10.65.20.88 tem-se que o segundo número mais à esquerda, ou seja, 65, é o número da quadra. No exemplo dado, a cartografia apresentada especifica a 65^a quadra em que o distrito foi subdividido.

Regra de aprendizado

É uma metodologia descrita por um algoritmo através do qual a Rede Neural adquire conhecimento durante a fase de treinamento.

As regras de aprendizado podem ser de dois tipos principais: a regra de Hebb é utilizada no treinamento não supervisionado, ou seja, quando apenas a entrada é conhecida; já a regra Delta ou regra dos erros mínimos quadrados é utilizada no treinamento supervisionado, quando tanto a entrada quanto a saída desejada são conhecidas. Outras regras podem ser desenvolvidas a partir das duas mencionadas e de acordo com a aplicabilidade no caso prático orientado pelo Operador da Rede.

Sinapse

É o ponto de contato do axônio de um neurônio com o dendrito de outro. É pela sinapse que fica controlado o fluxo de informação entre os neurônios, viabilizando o aprendizado da Rede Neural.

A sinapse libera elementos químicos chamados de neurotransmissores sempre que o axônio emite quantidade suficiente de sinais em um curto espaço de tempo, gerando a ativação ou o disparo da célula, que transmite seus impulsos para a célula seguinte. A força ou potencial de ativação capaz de disparar o “conhecimento” para outra célula é o grande responsável pelo aprendizado da rede, força esta denominada de peso e ajustada ao longo do treinamento.

Soma

É o corpo da célula, ou seja, do neurônio, medindo apenas alguns milésimos de milímetros segundo Braga, Carvalho e Ludemir (2000).

Supertreinamento

Situação em que o aprendizado da rede ultrapassa os limites ótimos, tornando a rede “viciada” apenas no conjunto de dados submetido ao treinamento.

Acontece em virtude da escolha inadequada dos critérios de parada. Quando ocorre o supertreinamento, ao ser apresentado à rede um dado que não foi utilizado para o treinamento, a probabilidade de que seja classificado incorretamente é potencializada.

Taxa de aprendizagem (α)

É um valor, entre zero e um, utilizado para dar relevância ao ajuste de pesos do neurônio vencedor e de seus vizinhos.

A taxa de aprendizagem é inicializada com um valor entre zero e um, preferencialmente próximo de 1 (tal como 0,9), sendo diminuída com o decorrer do tempo, a fim de proporcionar maior especialização à rede, fazendo uma espécie de “ajuste fino” com os pesos.

Tolerância a falhas

É a habilidade que possui a estrutura cerebral de prosseguir com a consecução de uma tarefa mesmo depois da perda parcial ou definitiva de um neurônio.

A expressiva tolerância a falhas é consequência da distribuição do trabalho executado em paralelo pelos neurônios cerebrais, conhecidos por realizarem um *processamento distribuído*.

Topologia da vizinhança

Indica as direções em que será projetada a influência do neurônio vencedor, definindo a velocidade de especialização do aprendizado da RNA. É constante ao longo de todo o treinamento.

As topologias mais comuns são a quadrada, em cruz e em estrela, mas outros formatos podem ser definidos para treinamento da rede.

A topologia em cruz define maior especialização, por influenciar menos neurônios. A quadrada define a menor especialização de todas por influenciar maior quantidade de neurônios. A topologia em estrela tem nível de especialização intermediário entre a topologia em cruz e a topologia quadrada.

Treinamento

É o procedimento, através do qual, a rede adquire conhecimento a partir da análise de similaridade entre as entradas e os neurônios da Camada de Saída (ou de Competição). É sinônimo de aprendizado.

Antes que tenha início, o Operador da Rede deve estabelecer parâmetros de treinamento tais como a quantidade de neurônios da Camada de Saída, os critérios de parada, a topologia e o raio da vizinhança e outros.

Ao ser iniciado o treinamento, a rede que identificará qual dos neurônios da Camada de Saída tem maior similaridade com a entrada apresentada. A verificação da similaridade é feita através do cálculo da distância entre a entrada e cada neurônio da Camada de Saída, descobrindo, assim, qual dos neurônios desta camada possui menor distância – e conseqüentemente maior proximidade – em relação à entrada. O neurônio vencedor, ou seja, o mais próximo da entrada tem seu peso incrementado, tanto quanto os seus neurônios vizinhos. Através deste ajuste de pesos, o aprendizado vai sendo implementado. Quando a rede alcança um dos critérios de parada, o aprendizado é terminado e a rede é considerada treinada, restando definidos os pesos finais entre as camadas de Entrada e de Saída, para fins de classificação de qualquer nova entrada apresentada à rede.

APÊNDICES

Apêndice A – Listagem da pesquisa de composição imobiliária

DISTRITO 15

Quantitativo de imóveis por quadra, conforme o atributo **ÁREA** DO TERRENO > 1000.

Quadra	1 :	12	imóveis.
Quadra	2 :	111	imóveis.
Quadra	3 :	1	imóvel.
Quadra	4 :	58	imóveis.
Quadra	5 :	41	imóveis.
Quadra	6 :	2	imóveis.
Quadra	7 :	139	imóveis.
Quadra	8 :	105	imóveis.
Quadra	9 :	75	imóveis.
Quadra	10 :	20	imóveis.
Quadra	12 :	22	imóveis.
Quadra	13 :	74	imóveis.
Quadra	15 :	86	imóveis.
Quadra	16 :	66	imóveis.
Quadra	17 :	2	imóveis.
Quadra	18 :	90	imóveis.
Quadra	19 :	58	imóveis.
Quadra	21 :	74	imóveis.
Quadra	22 :	87	imóveis.
Quadra	23 :	1	imóvel.
Quadra	24 :	2	imóveis.
Quadra	25 :	35	imóveis.
Quadra	26 :	163	imóveis.
Quadra	27 :	66	imóveis.
Quadra	28 :	43	imóveis.
Quadra	29 :	3	imóveis.
Quadra	30 :	45	imóveis.
Quadra	31 :	32	imóveis.
Quadra	32 :	122	imóveis.
Quadra	35 :	4	imóveis.
Quadra	36 :	105	imóveis.
Quadra	37 :	120	imóveis.
Quadra	38 :	91	imóveis.
Quadra	39 :	146	imóveis.
Quadra	40 :	30	imóveis.
Quadra	41 :	34	imóveis.
Quadra	42 :	3	imóveis.
Quadra	44 :	2	imóveis.
Quadra	45 :	20	imóveis.
Quadra	46 :	94	imóveis.
Quadra	48 :	22	imóveis.
Quadra	49 :	19	imóveis.
Quadra	51 :	54	imóveis.
Quadra	52 :	4	imóveis.
Quadra	53 :	11	imóveis.
Quadra	54 :	6	imóveis.
Quadra	55 :	4	imóveis.
Quadra	56 :	2	imóveis.
Quadra	57 :	2	imóveis.
Quadra	58 :	6	imóveis.
Quadra	59 :	2	imóveis.
Quadra	60 :	2	imóveis.

Quadra	61 :	67	imóveis.
Quadra	62 :	67	imóveis.
Quadra	63 :	23	imóveis.
Quadra	64 :	81	imóveis.
Quadra	65 :	70	imóveis.
Quadra	66 :	1	imóvel.
Quadra	67 :	193	imóveis.
Quadra	68 :	202	imóveis.
Quadra	69 :	1	imóvel.
Quadra	70 :	16	imóveis.
Quadra	72 :	26	imóveis.
Quadra	73 :	61	imóveis.
Quadra	74 :	74	imóveis.
Quadra	75 :	299	imóveis.
Quadra	76 :	6	imóveis.
Quadra	77 :	2	imóveis.
Quadra	78 :	2	imóveis.
Quadra	79 :	2	imóveis.
Quadra	80 :	53	imóveis.
Quadra	81 :	24	imóveis.
Quadra	82 :	3	imóveis.
Quadra	83 :	3	imóveis.
Quadra	84 :	41	imóveis.
Quadra	86 :	2	imóveis.
Quadra	87 :	20	imóveis.
Quadra	88 :	93	imóveis.
Quadra	89 :	50	imóveis.
Quadra	90 :	1	imóvel.
Quadra	91 :	22	imóveis.
Quadra	92 :	41	imóveis.
Quadra	93 :	103	imóveis.
Quadra	94 :	25	imóveis.
Quadra	95 :	65	imóveis.
Quadra	96 :	60	imóveis.
Quadra	97 :	13	imóveis.
Quadra	98 :	2	imóveis.
Quadra	99 :	15	imóveis.
Quadra	101 :	1	imóvel.
Quadra	102 :	1	imóvel.
Quadra	103 :	4	imóveis.
Quadra	104 :	1	imóvel.
Quadra	105 :	2	imóveis.
Quadra	106 :	98	imóveis.
Quadra	109 :	1	imóvel.
Quadra	111 :	1	imóvel.
Quadra	113 :	45	imóveis.
Quadra	114 :	34	imóveis.
Quadra	116 :	91	imóveis.
Quadra	118 :	20	imóveis.
Quadra	119 :	17	imóveis.
Quadra	120 :	24	imóveis.
Quadra	121 :	19	imóveis.
Quadra	122 :	2	imóveis.
Quadra	123 :	47	imóveis.
Quadra	127 :	24	imóveis.
Quadra	128 :	2	imóveis.
Quadra	129 :	122	imóveis.
Quadra	130 :	322	imóveis.

Quadra	131	:	21	imóveis.
Quadra	132	:	504	imóveis.
Quadra	133	:	98	imóveis.
Quadra	134	:	130	imóveis.
Quadra	135	:	543	imóveis.
Quadra	136	:	5	imóveis.
Quadra	137	:	5	imóveis.
Quadra	138	:	316	imóveis.
Quadra	139	:	17	imóveis.
Quadra	140	:	62	imóveis.
Quadra	141	:	75	imóveis.
Quadra	142	:	6	imóveis.
Quadra	143	:	1	imóvel.

**Quantitativo DECRESCENTE de
imóveis por quadra, conforme o
atributo ÁREA DO TERRENO > 1000.**

Quadra	135	:	543	imóveis.
Quadra	132	:	504	imóveis.
Quadra	130	:	322	imóveis.
Quadra	138	:	316	imóveis.
Quadra	75	:	299	imóveis.
Quadra	68	:	202	imóveis.
Quadra	67	:	193	imóveis.
Quadra	26	:	163	imóveis.
Quadra	39	:	146	imóveis.
Quadra	7	:	139	imóveis.
Quadra	134	:	130	imóveis.
Quadra	32	:	122	imóveis.
Quadra	129	:	122	imóveis.
Quadra	37	:	120	imóveis.
Quadra	2	:	111	imóveis.
Quadra	8	:	105	imóveis.
Quadra	36	:	105	imóveis.
Quadra	93	:	103	imóveis.
Quadra	106	:	98	imóveis.
Quadra	133	:	98	imóveis.
Quadra	46	:	94	imóveis.
Quadra	88	:	93	imóveis.
Quadra	38	:	91	imóveis.
Quadra	116	:	91	imóveis.
Quadra	18	:	90	imóveis.
Quadra	22	:	87	imóveis.
Quadra	15	:	86	imóveis.
Quadra	64	:	81	imóveis.
Quadra	9	:	75	imóveis.
Quadra	141	:	75	imóveis.
Quadra	13	:	74	imóveis.
Quadra	21	:	74	imóveis.
Quadra	74	:	74	imóveis.
Quadra	65	:	70	imóveis.
Quadra	61	:	67	imóveis.
Quadra	62	:	67	imóveis.
Quadra	16	:	66	imóveis.
Quadra	27	:	66	imóveis.
Quadra	95	:	65	imóveis.
Quadra	140	:	62	imóveis.
Quadra	73	:	61	imóveis.
Quadra	96	:	60	imóveis.
Quadra	4	:	58	imóveis.

Quadra	19	:	58	imóveis.
Quadra	51	:	54	imóveis.
Quadra	80	:	53	imóveis.
Quadra	89	:	50	imóveis.
Quadra	123	:	47	imóveis.
Quadra	30	:	45	imóveis.
Quadra	113	:	45	imóveis.
Quadra	28	:	43	imóveis.
Quadra	5	:	41	imóveis.
Quadra	84	:	41	imóveis.
Quadra	92	:	41	imóveis.
Quadra	25	:	35	imóveis.
Quadra	41	:	34	imóveis.
Quadra	114	:	34	imóveis.
Quadra	31	:	32	imóveis.
Quadra	40	:	30	imóveis.
Quadra	72	:	26	imóveis.
Quadra	94	:	25	imóveis.
Quadra	81	:	24	imóveis.
Quadra	120	:	24	imóveis.
Quadra	127	:	24	imóveis.
Quadra	63	:	23	imóveis.
Quadra	12	:	22	imóveis.
Quadra	48	:	22	imóveis.
Quadra	91	:	22	imóveis.
Quadra	131	:	21	imóveis.
Quadra	10	:	20	imóveis.
Quadra	45	:	20	imóveis.
Quadra	87	:	20	imóveis.
Quadra	118	:	20	imóveis.
Quadra	49	:	19	imóveis.
Quadra	121	:	19	imóveis.
Quadra	119	:	17	imóveis.
Quadra	139	:	17	imóveis.
Quadra	70	:	16	imóveis.
Quadra	99	:	15	imóveis.
Quadra	97	:	13	imóveis.
Quadra	1	:	12	imóveis.
Quadra	53	:	11	imóveis.
Quadra	54	:	6	imóveis.
Quadra	58	:	6	imóveis.
Quadra	76	:	6	imóveis.
Quadra	142	:	6	imóveis.
Quadra	136	:	5	imóveis.
Quadra	137	:	5	imóveis.
Quadra	35	:	4	imóveis.
Quadra	52	:	4	imóveis.
Quadra	55	:	4	imóveis.
Quadra	103	:	4	imóveis.
Quadra	29	:	3	imóveis.
Quadra	42	:	3	imóveis.
Quadra	82	:	3	imóveis.
Quadra	83	:	3	imóveis.
Quadra	6	:	2	imóveis.
Quadra	17	:	2	imóveis.
Quadra	24	:	2	imóveis.
Quadra	44	:	2	imóveis.
Quadra	56	:	2	imóveis.
Quadra	57	:	2	imóveis.
Quadra	59	:	2	imóveis.
Quadra	60	:	2	imóveis.

Quadra	77 :	2 imóveis.	Quadra	66 :	1 imóvel.
Quadra	78 :	2 imóveis.	Quadra	69 :	1 imóvel.
Quadra	79 :	2 imóveis.	Quadra	90 :	1 imóvel.
Quadra	86 :	2 imóveis.	Quadra	101 :	1 imóvel.
Quadra	98 :	2 imóveis.	Quadra	102 :	1 imóvel.
Quadra	105 :	2 imóveis.	Quadra	104 :	1 imóvel.
Quadra	122 :	2 imóveis.	Quadra	109 :	1 imóvel.
Quadra	128 :	2 imóveis.	Quadra	111 :	1 imóvel.
Quadra	3 :	1 imóvel.	Quadra	143 :	1 imóvel.
Quadra	23 :	1 imóvel.			

Apêndice B – Treinamento (ajuste de pesos) da RNA

Analizador do Solo Urbano vr 1.4.1

Copyright (C) 2007-2010, Carlos Cristiano Cabral - Fortaleza-Ce, BRASIL

Relatório gerado em 18/05/2010 às 23:42:30.

APRENDIZADO NÃO SUPERVISIONADO - Processado sobre o distrito 15

=====

NEURÔNIOS DA CAMADA DE ENTRADA

1. Área edificada
2. Tipo do imóvel

ORDEM DA CAMADA DE SAÍDA

Camada de saída (competição) montada em matriz de ordem 2x2 (4 neurônios)

TOPOLOGIA DA VIZINHANÇA

A topologia de atualização da vizinhança do neurônio vencedor é EM ESTRELA

RAIO DA VIZINHANÇA

O raio inicial da vizinhança é de 5 unidade(s).

1ª E N T R A D A

CALCULANDO DISTÂNCIAS

Analizando o neurônio 1x1

Cálculo de distâncias

- (A) 0,2500 = Valor normalizado do neurônio de entrada ÁREA EDIFICADA
- (B) 0,0630 = Peso do neurônio 1x1 em relação à entrada
- (C) 0,0349 = Fator parcial calculado pela fórmula $C=(A-B)^2$

- (A) 0,5000 = Valor normalizado do neurônio de entrada TIPO DO IMÓVEL
- (B) 0,1598 = Peso do neurônio 1x1 em relação à entrada
- (C) 0,1157 = Fator parcial calculado pela fórmula $C=(A-B)^2$

- (D) 0,1506 = Distância calculada pelo somatório dos fatores parciais
(Somadas dos C's)

Analizando o neurônio 1x2

Cálculo de distâncias

- (A) 0,2500 = Valor normalizado do neurônio de entrada ÁREA EDIFICADA
- (B) 0,0221 = Peso do neurônio 1x2 em relação à entrada
- (C) 0,0519 = Fator parcial calculado pela fórmula $C=(A-B)^2$

- (A) 0,5000 = Valor normalizado do neurônio de entrada TIPO DO IMÓVEL
- (B) 0,2340 = Peso do neurônio 1x2 em relação à entrada
- (C) 0,0707 = Fator parcial calculado pela fórmula $C=(A-B)^2$

- (D) 0,1226 = Distância calculada pelo somatório dos fatores parciais
(Somadas dos C's)

Analisando o neurônio 2x1

Cálculo de distâncias

- (A) 0,2500 = Valor normalizado do neurônio de entrada ÁREA EDIFICADA
- (B) 0,1750 = Peso do neurônio 2x1 em relação à entrada
- (C) 0,0056 = Fator parcial calculado pela fórmula $C=(A-B)^2$

- (A) 0,5000 = Valor normalizado do neurônio de entrada TIPO DO IMÓVEL
- (B) 0,0541 = Peso do neurônio 2x1 em relação à entrada
- (C) 0,1988 = Fator parcial calculado pela fórmula $C=(A-B)^2$

- (D) 0,2044 = Distância calculada pelo somatório dos fatores parciais (Somados dos C's)

Analisando o neurônio 2x2

Cálculo de distâncias

- (A) 0,2500 = Valor normalizado do neurônio de entrada ÁREA EDIFICADA
- (B) 0,1528 = Peso do neurônio 2x2 em relação à entrada
- (C) 0,0094 = Fator parcial calculado pela fórmula $C=(A-B)^2$

- (A) 0,5000 = Valor normalizado do neurônio de entrada TIPO DO IMÓVEL
- (B) 0,0225 = Peso do neurônio 2x2 em relação à entrada
- (C) 0,2280 = Fator parcial calculado pela fórmula $C=(A-B)^2$

- (D) 0,2374 = Distância calculada pelo somatório dos fatores parciais (Somados dos C's)

DETERMINANDO NEURÔNIO VENCEDOR

Neurônio 1x2 venceu, pois possuiu a menor distância calculada: 0,1226

AJUSTE DE PESO DO NEURÔNIO VENCEDOR

Ajustando peso do neurônio 1x2

- (A) 0,2500 = Valor normalizado do neurônio de entrada ÁREA EDIFICADA
- (B) 0,0221 = Peso do neurônio 1x2 em relação à entrada
- (C) 0,9000 = Alfa

- (D) 0,2272 = Novo peso em relação à entrada ÁREA EDIFICADA ajustado pela fórmula $D=B+C(A-B)$

- (A) 0,5000 = Valor normalizado do neurônio de entrada TIPO DO IMÓVEL
- (B) 0,2340 = Peso do neurônio 1x2 em relação à entrada
- (C) 0,9000 = Alfa

- (D) 0,4734 = Novo peso em relação à entrada TIPO DO IMÓVEL ajustado pela fórmula $D=B+C(A-B)$

AJUSTA PESOS DOS VIZINHOS na matriz 2x2

Ajustando peso do neurônio 2x2

- (A) 0,2500 = Valor normalizado do neurônio de entrada ÁREA EDIFICADA
- (B) 0,1528 = Peso do neurônio 2x2 em relação à entrada
- (C) 0,9000 = Alfa

(D) 0,2402 = Novo peso em relação à entrada ÁREA EDIFICADA ajustado pela fórmula $D=B+C(A-B)$

(A) 0,5000 = Valor normalizado do neurônio de entrada TIPO DO IMÓVEL

(B) 0,0225 = Peso do neurônio 2x2 em relação à entrada

(C) 0,9000 = Alfa

(D) 0,4522 = Novo peso em relação à entrada TIPO DO IMÓVEL ajustado pela fórmula $D=B+C(A-B)$

Ajustando peso do neurônio 2x1

(A) 0,2500 = Valor normalizado do neurônio de entrada ÁREA EDIFICADA

(B) 0,1750 = Peso do neurônio 2x1 em relação à entrada

(C) 0,9000 = Alfa

(D) 0,2425 = Novo peso em relação à entrada ÁREA EDIFICADA ajustado pela fórmula $D=B+C(A-B)$

(A) 0,5000 = Valor normalizado do neurônio de entrada TIPO DO IMÓVEL

(B) 0,0541 = Peso do neurônio 2x1 em relação à entrada

(C) 0,9000 = Alfa

(D) 0,4554 = Novo peso em relação à entrada TIPO DO IMÓVEL ajustado pela fórmula $D=B+C(A-B)$

Ajustando peso do neurônio 1x1

(A) 0,2500 = Valor normalizado do neurônio de entrada ÁREA EDIFICADA

(B) 0,0630 = Peso do neurônio 1x1 em relação à entrada

(C) 0,9000 = Alfa

(D) 0,2313 = Novo peso em relação à entrada ÁREA EDIFICADA ajustado pela fórmula $D=B+C(A-B)$

(A) 0,5000 = Valor normalizado do neurônio de entrada TIPO DO IMÓVEL

(B) 0,1599 = Peso do neurônio 1x1 em relação à entrada

(C) 0,9000 = Alfa

(D) 0,4659 = Novo peso em relação à entrada TIPO DO IMÓVEL ajustado pela fórmula $D=B+C(A-B)$

ANEXOS

Anexo A – Tabela de conversão de distrito para bairro

Distrito	Bairro	Distrito	Bairro	Distrito	Bairro
1	Centro	37	Bela Vista	61 - Parte 01	Cidade dos Funcionários
2	Centro	38	Montese	61 - Parte 02	Cidade dos Funcionários
3	Centro	39	Itaoca	62 - Parte 01	Cidade dos Funcionários
4	Moura Brasil	40 - Parte 01	Aeroporto	62 - Parte 02	Castelão
5	Praia de Iracema	40 - Parte 02	Dias Macêdo	62 - Parte 03	Castelão
6	Benfica - Farias Brito	40 - Parte 03	Aeroporto	63 - Parte 01	Barroso
7	Aldeota	41	Parangaba	63 - Parte 02	Cajazeiras
8	Parque Araxá - Monte Castelo	42	Vila Peri	64 - Parte 01	Lagoa Redonda
9	Meireles	43	Maraponga	64 - Parte 02	Lagoa Redonda
10	Rodolfo Teófilo	44 - Parte 01	Mondubim	64 - Parte 03	Lagoa Redonda
11	José Bonifácio	44 - Parte 02	Dendê	64 - Parte 04	Lagoa Redonda
12	Benfica	45	Seminha	65 - Parte 01	Paupina
13	Joaquim Távora	46	Aerolândia	65 - Parte 02	Paupina
14 - Parte 01	Bairro de Fátima	47 - Parte 01	Cidade 2000	65 - Parte 03	Paupina
14 - Parte 02	Vila União	47 - Parte 02	Papicu - Cidade 2000	65 - Parte 04	Paupina
15	Aldeota	48	Varjota - Mucunipe	65 - Parte 05	Coaçú
16	Dionísio Torres	49	Mucunipe	66 - Parte 01	Prefeito José Walter
17	Tauape	50 - Parte 01	Siqueira	66 - Parte 02	Prefeito José Walter
18	Aldeota	50 - Parte 02	Siqueira	66 - Parte 03	Prefeito José Walter
19	Varjota	50 - Parte 03	Siqueira	67	Prefeito José Walter
20	Vila Ellery	51	Papicu	68 - Parte 01	Passaré
21	Jaracacanga	52	Praia do Futuro	68 - Parte 02	Parque 2 Irmãos
22 - Parte 01	Bairro do Ceará	53	Praia do Futuro	68 - Parte 03	Passaré
22 - Parte 02	Bairro do Ceará	54	Praia do Futuro	68 - Parte 04	Parque 2 Irmãos
23 - Parte 01	Alvaro Weyne	55 - Parte 01	Cocó	69 - Parte 01	Pedras
23 - Parte 02	Bairro do Ceará	55 - Parte 02	Cocó	69 - Parte 02	Ancuri
24	Vila Velha	56 - Parte 01	Salinas	69 - Parte 03	Ancuri
25	Jardim Guanabara	56 - Parte 02	Guarapes	70 - Parte 01	Conjunto Ceará
26 - Parte 01	São Gerardo	56 - Parte 03	Engenheiro Luciano Cavalcante	70 - Parte 02	Conjunto Ceará
26 - Parte 02	Presidente Kennedy	56 - Parte 04	Engenheiro Luciano Cavalcante	71 - Parte 01	Mondubim
27	Parquelândia	56 - Parte 05	Jardim das Oliveiras	71 - Parte 02	Parque Santa Rosa
28 - Parte 01	Pici	57 - Parte 01	Edson Queiroz	71 - Parte 03	Canindezinho
28 - Parte 02	Pici	57 - Parte 02	Edson Queiroz	71 - Parte 04	Canindezinho
29 - Parte 01	Quintino Cunha	57 - Parte 03	Edson Queiroz	71 - Parte 05	Manoel Sátiro
29 - Parte 02	Antônio Bezerra	57 - Parte 04	Edson Queiroz	72 - Parte 01	Mondubim
30 - Parte 01	Dom Lustosa	57 - Parte 05	Edson Queiroz	72 - Parte 02	Mondubim
30 - Parte 02	Henrique Jorge	58 - Parte 01	Sabiaguaba	73	Genibaú
31	Granja Portugal	58 - Parte 02	Sabiaguaba	74 - Parte 01	Jangurussú
32 - Parte 01	Granja Lisboa	58 - Parte 03	Sabiaguaba	74 - Parte 02	Jangurussú
32 - Parte 02	Granja Portugal	58 - Parte 04	Sabiaguaba	74 - Parte 03	Jangurussú
33 - Parte 01	Bom Jardim	59 - Parte 01	Alagadiço Novo	75 - Parte 01	Jangurussú
33 - Parte 02	Bom Jardim	59 - Parte 02	Sapiranga - Coité	75 - Parte 02	Jangurussú
34	Bom Sucesso	59 - Parte 03	Sapiranga - Coité	76 - Parte 01	Messejana
35	Henrique Jorge	60 - Parte 01	Cambeba	76 - Parte 02	Messejana
36	Jôquei Clube	60 - Parte 02	Parque Iracema	77	Aldeota

Anexo B – Cálculo do IPTU

Lei nº 8.703 de 30 de abril de 2003 (Anexo IV)⁵
 FÓRMULA DE CÁLCULO DO VALOR VENAL – IPTU

$$Vv = \{[0.8415814466 \times (\sqrt[64]{At \times Tp \times Tp}) \times At] \times Fi \times Vt \times FI\} + (Au \times Ve \times Fd)$$

Vv	– Valor venal (em R\$)	At	– Área do terreno (em m ²)
Vt	– Valor do metro quadrado do terreno (em R\$)	Tp	– Testada principal (em m)
FI	– Fator correção do lote	Ae	– Área total edificada (em m ²)
Au	– Área edificada da unidade (em m ²)	Fd	– Fator de depreciação
Ve ⁶	– Valor do metro quadrado da edificação (em R\$)	Fi	– Fração Ideal

Se $Au > 0$ e $Ae > 0$, $Fi = (Au/Ae)$

1. Fator de Depreciação (Fd)

Faixa em anos	Fator aplicado (%)
Menor ou igual a 5	0
Maior que 5 e menor ou igual a 10	7
Maior que 10 e menor ou igual a 15	14
Maior que 15 e menor ou igual a 20	21
Maior que 20 e menor ou igual a 25	28
Maior que 25 e menor ou igual a 30	35
Maior que 30 e menor ou igual a 35	42
Maior que 35 e menor ou igual a 40	49
Maior que 40 e menor ou igual a 45	56
Maior que 45 e menor ou igual a 50	63
Acima de 50	70

2. Fator de Correção do Lote (FI = soma dos pesos/12)

PESOS:

SITUAÇÃO DO LOTE:

Normal	1,00
Esquina	1,20
Vila	0,75
Encravado	0,50
Quadra	1,80
Gleba	2,20

TOPOGRAFIA:

Plano	1,00
Aclive	0,80
Declive	0,75
Topografia irregular	0,85

⁵ Formatação nossa.

⁶ O valor da edificação (Ve) é obtido enquadrando-se o fator de edificação (FE_d) calculado, na tabela de valores de edificação constante do anexo II da lei 8703/2003 (vide Anexo C).

PEDOLOGIA:

Arenoso	1,00
Argiloso.....	1,00
Rochoso	0,40
Arenoso e argiloso	1,00
Arenoso e rochoso	0,60
Argiloso e rochoso.....	0,60
Inundável.....	0,40
Pantanosos	0,30
Sujeito a marés.....	0,40

OCUPAÇÃO:

Sem	1,00
Em construção.....	1,00
Construção paralisada	1,00
Ruínas ou demolição.....	1,00
Edificado.....	1,00
Estacionamento.....	1,20
Laser.....	0,50
Agricultura	0,50
Depósito	1,30

PAVIMENTAÇÃO PARA PEDESTRE:

Sem	0,50
Tijoleira	0,80
Pedra rústica	0,90
Cimentado	1,00
Pré-moldado.....	1,05
Mosaico	1,10
Pedra portuguesa.....	1,20

PAVIMENTAÇÃO PARA VEÍCULOS:

Sem	0,50
Pedra rústica	1,00
Paralelepípedo	1,05
Pré-moldado.....	1,10
Concreto	1,20
Asfalto.....	1,15

REDE DE ÁGUA:

Sem	0,50
Com	1,00

REDE DE ESGOTO:

Sem	0,60
Com	1,00

GALERIA PLUVIAL:

Sem	0,50
Com	1,00

GUIAS/SARJETAS:

Sem	0,50
Com	1,00

ILUMINAÇÃO PÚBLICA:

Sem	0,30
Com	1,00

ARBORIZAÇÃO:

Sem	1,00
Com	0,50

3. Fator de Edificação (FE_d = soma dos pesos/17)**PESOS:****SITUAÇÃO RELATIVA AO LOGRADOURO:**

Frente	1,00
Fundo	0,80
Galeria.....	0,90
Vila.....	0,80

SITUAÇÃO RELATIVA AO LOTE:

Isolado recuado	1,00
Isolado alinhado	1,00
Rec. s/ espaço lateral.....	1,00
Isolado superposto	1,00
Isolado superposto alinhado	1,00
Superposto s/ espaço lateral recuado....	1,00

CLASSIFICAÇÃO ARQUITETÔNICA:

Choça ou barraco	0,00
Casa	1,00

Apartamento.....	2,00
Apartamento de cobertura.....	3,00
Sala	2,00
Conjunto de salas.....	2,50
Loja.....	1,50
Sobreloja	2,00
Subsolo.....	0,80
Galpão fechado	0,90
Galpão aberto.....	0,80
Estacionamento coberto.....	0,80
Arquitetura especial.....	2,00

COBERTURA:

Palha	0,00
Cerâmica	1,00
Fibra-cimento.....	1,10
Laje.....	1,15
Metálica	1,25

Especial 1,35

PISO:

Terra batida 0,00
 Tijoleira 0,50
 Cimento 0,70
 Mosaico 0,80
 Cerâmica 1,00
 Madeira taco comum 1,00
 Madeira taco parquet 1,50
 Madeira assoalho 1,80
 Plástico 1,20
 Pedra 1,90
 Marmorite 1,20
 Mármore 2,10
 Especial 2,20

FORRO:

Sem 0,00
 Madeira 0,85
 Estuque 0,90
 Laje 1,00
 Especial 1,50

ACABAMENTO INTERNO:

Sem 0,00
 Caiação 0,60
 Pintura impermeável 0,90
 Pintura a óleo 0,95
 Pintura plástica 1,10
 Aparente rústico 1,00
 Aparente luxo 1,50
 Especial 1,80

ACABAMENTO EXTERNO:

Sem 0,00
 Caiação 0,60
 Pintura impermeável 0,90
 Pintura a óleo 0,95
 Pintura plástica 1,10
 Aparente rústico 1,00
 Aparente luxo 1,50
 Especial 1,80

ESQUADRIAS:

Sem 0,00
 Madeira 1,00
 Rústica 1,10
 Ferro 1,20

Alumínio 1,30
 Mista 1,40
 Especial 1,50

VIDROS:

Sem 0,50
 Comum 0,90
 Vitrais 1,10
 Fumê 1,30
 Misto 1,00
 Especiais 1,50

INSTALAÇÃO ELÉTRICA NA UNIDADE:

Sem 0,00
 Embutida 1,00
 Semi-embutida 0,80
 Aparente simples 0,50
 Aparente luxo 1,20

INSTALAÇÃO SANITÁRIA NA UNIDADE:

Sem 0,00
 Externa com fossa 0,50
 Externa com rede 0,60
 Interna com fossa simples 0,70
 Interna com fossa completa-rede 0,90
 Interna + 1 c/ fossa 1,00
 Interna + 1 c/ rede 1,10

CONSERVAÇÃO:

Boa 1,00
 Regular 0,90
 Má 0,50

OUTROS:

Não registrada 1,00
 Com jardim 1,20

OUTROS:

Não registrada 1,00
 Pomar ou horta 1,20

OUTROS:

Não registrada 1,00
 Piscina 2,20

OUTROS:

Não registrada 1,00
 Garagem 1,20

Anexo C – Tabela de valores de edificação

ANEXO II Tabela de Valor Unitário de M² de Edificação para 2004 (Anexo II da Lei nº 8.703 de 30.04.2003) (CLASSIFICAÇÃO ARQUITETÔNICA - TIPO - PADRÃO E VALORES)

VARIACÃO DO FATOR - EDIFICAÇÃO	CASA POPULAR TIPO 1	CASA TIPO 2	APTO TIPO 3	APTO COBERT TIPO 4	SALA TIPO 5	CONJ. DE SALA TIPO 6	LOJA TIPO 7	SOBRE LOJA TIPO 8	SUB SOLO TIPO 9	GALPÃO FECHADO TIPO 10	GALPÃO ABERTO TIPO 11	ESTACIONAMENTO COBERTO TIPO 12	ARQUITETURA ESPECÍFICA TIPO 13	CLASSE
PADRÃO LUXO														
> 1,39	337,28	605,37	529,82	529,82	352,25	352,25	457,95	352,25	190,22	246,58	122,57	190,22	529,82	12
1,31 A 1,39	270,53	485,56	422,71	422,71	281,81	281,81	366,35	281,81	152,16	197,27	98,63	152,16	422,71	11
1,22 A 1,30	216,66	388,87	339,58	339,58	225,44	225,44	293,08	225,44	122,57	157,81	78,90	122,57	339,58	10
PADRÃO ALTO NÍVEL														
1,13 A 1,21	173,32	311,09	271,94	271,94	180,35	180,35	233,89	180,35	97,22	126,81	61,98	97,22	271,94	9
1,04 A 1,12	138,18	248,02	216,98	216,98	145,13	145,13	187,39	145,13	77,48	101,43	50,71	77,48	216,98	8
0,96 A 1,03	111,24	199,66	140,89	173,31	115,53	115,53	150,76	115,53	61,98	80,31	39,43	61,98	173,31	7
PADRÃO NORMAL														
0,86 A 0,94	88,98	126,09	112,72	139,48	92,98	92,98	119,76	92,98	49,31	64,81	32,40	49,31	139,48	6
0,77 A 0,85	70,25	102,97	90,17	111,31	74,67	74,67	96,81	74,67	39,43	52,11	25,35	39,43	111,31	5
0,68 A 0,76	56,20	79,85	71,85	88,76	59,17	59,17	78,07	59,17	32,40	40,85	21,11	32,40	88,76	4
PADRÃO BAIXO NÍVEL														
0,59 A 0,67	45,65	63,04	46,45	70,43	46,48	46,48	61,98	46,48	25,35	33,80	16,89	25,35	70,43	3
0,50 A 0,58	36,29	50,42	36,61	56,35	36,61	36,61	49,31	36,61	21,11	26,76	12,67	21,11	56,35	2
ATÉ 0,49	29,26	39,92	29,57	45,07	29,57	29,57	39,43	29,57	16,89	21,11	9,85	16,89	45,07	1