



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2014/02.28.18.48-TDI

**CLASSIFICAÇÃO DE PADRÕES TEMPORAIS DE USO
DO SOLO E COBERTURA DA TERRA EM SÉRIES
TEMPORAIS DE ÍNDICE DE VEGETAÇÃO
UTILIZANDO UM SISTEMA NEURO-DIFUSO**

Toni Roberto Gomes Pimentel

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Sandra Aparecida Sandri, e Fernando Manuel Ramos, aprovada em 17 de abril de 2014.

URL do documento original:

<<http://urlib.net/8JMKD3MGP5W34M/3FRC4FL>>

INPE
São José dos Campos
2014

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@sid.inpe.br

CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO DA PRODUÇÃO INTELLECTUAL DO INPE (RE/DIR-204):**Presidente:**

Marciana Leite Ribeiro - Serviço de Informação e Documentação (SID)

Membros:

Dr. Gerald Jean Francis Banon - Coordenação Observação da Terra (OBT)

Dr. Amauri Silva Montes - Coordenação Engenharia e Tecnologia Espaciais (ETE)

Dr. André de Castro Milone - Coordenação Ciências Espaciais e Atmosféricas (CEA)

Dr. Joaquim José Barroso de Castro - Centro de Tecnologias Espaciais (CTE)

Dr. Manoel Alonso Gan - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Dr^a Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon - Coordenação de Observação da Terra (OBT)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Maria Tereza Smith de Brito - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

EDITORAÇÃO ELETRÔNICA:

Maria Tereza Smith de Brito - Serviço de Informação e Documentação (SID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2014/02.28.18.48-TDI

**CLASSIFICAÇÃO DE PADRÕES TEMPORAIS DE USO
DO SOLO E COBERTURA DA TERRA EM SÉRIES
TEMPORAIS DE ÍNDICE DE VEGETAÇÃO
UTILIZANDO UM SISTEMA NEURO-DIFUSO**

Toni Roberto Gomes Pimentel

Dissertação de Mestrado do Curso de Pós-Graduação em Computação Aplicada, orientada pelos Drs. Sandra Aparecida Sandri, e Fernando Manuel Ramos, aprovada em 17 de abril de 2014.

URL do documento original:

<http://urlib.net/8JMKD3MGP5W34M/3FRC4FL>

INPE
São José dos Campos
2014

Dados Internacionais de Catalogação na Publicação (CIP)

Pimentel, Toni Roberto Gomes.

P649c Classificação de padrões temporais de uso do solo e cobertura da terra em séries temporais de índice de vegetação utilizando um sistema neuro-difuso / Toni Roberto Gomes Pimentel. – São José dos Campos : INPE, 2014.

xx + 82 p. ; (sid.inpe.br/mtc-m21b/2014/02.28.18.48-TDI)

Dissertação (Mestrado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2014.

Orientadores : Drs. Sandra Aparecida Sandri, e Fernando Manuel Ramos.

1. Índice de vegetação. 2. Classificação de padrões. 3. Uso do solo e cobertura da terra. 4. Sistema neuro-difuso. 5. Time series. I.Título.

CDU 519.246.8



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de **Mestre** em
Computação Aplicada

Dr. Haroldo Fraga de Campos Velho



Presidente / INPE / São José dos Campos - SP

Dra. Sandra Aparecida Sandri



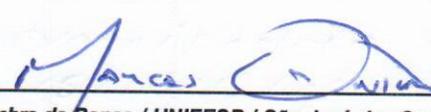
Orientador(a) / INPE / SJC Campos - SP

Dr. Fernando Manuel Ramos



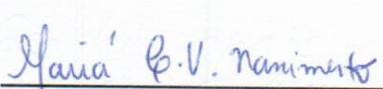
Orientador(a) / INPE / SJC Campos - SP

Dr. Marcos Gonçalves Quiles



Membro da Banca / UNIFESP / São José dos Campos - SP

Dra. Mariá Cristina Vasconcelos
Nascimento



Convidado(a) / UNIFESP / São José dos Campos - SP

Este trabalho foi aprovado por:

() maioria simples

() unanimidade

Aluno (a): **Toni Roberto Gomes Pimentel**

São José dos Campos, 17 de Abril de 2014

“Se você andar sabiamente, nada atrapalhará o seu caminho, e você não tropeçará quando correr”.

PROVÉRBIOS DE SALOMÃO

A Deus, minha esposa e minha família.

AGRADECIMENTOS

Agradeço a

Deus, por ser a essência da minha vida, o **Espírito** que me guia, o Autor da minha fé, o Socorro bem presente na hora da angústia;

Ingrid, minha esposa, e a toda minha família, que me deram muito carinho e apoio e não mediram esforços para me ajudar a passar por mais esta etapa;

Sandra Sandri e **Fernando Ramos**, meus professores e orientadores, pela paciência na orientação e pelo grande incentivo e ajuda, que tornaram possível a elaboração e conclusão deste trabalho;

todos os professores do curso de Computação Aplicada do INPE, que me passaram tanto conhecimento;

CNPQ, pela bolsa de mestrado concedida;

Jesus Cristo, meu Senhor e Salvador, que é o Alfa e o Omega, o Primeiro e o Último, o Princípio e o Fim, pois por meio dEle, Deus criou todas as coisas, no céu e na terra, tanto o que se vê como o que não se vê.

RESUMO

A questão das mudanças no uso do solo e cobertura da terra no Brasil tem despertado grande interesse, dentro e fora do meio científico, devido, principalmente, aos possíveis impactos ambientais e socioeconômicos que essas mudanças podem acarretar. Poluição do ar e da água, desertificação, desflorestamento ilegal e perda de biodiversidade são apenas alguns problemas relacionados a mudanças de uso do solo e cobertura da terra. Por isso, todas as precauções devem ser tomadas para que os ecossistemas locais e globais sejam minimamente afetados. Com base nesse interesse, se torna de suma importância o desenvolvimento de métodos, principalmente computacionais, capazes de monitorar e identificar mudanças de uso do solo e cobertura da terra no menor tempo possível, já que os processos de degradação do solo, em áreas de interesse público, são rápidos e, na maioria das vezes, devastadores. Os dados de sensoriamento remoto têm servido como base crucial para esse monitoramento, entretanto, normalmente as técnicas de monitoramento de uso do solo e cobertura da terra no Brasil demandam muito tempo e conhecimento especialista para serem realizadas, o que limita a possibilidade de ações conservacionistas pelas autoridades competentes. Essa foi a principal motivação do presente trabalho, e o objetivo traçado foi o de desenvolver um sistema capaz de classificar padrões temporais de uso do solo e cobertura da terra utilizando séries temporais de índice de vegetação. Utilizamos séries temporais obtidas a partir de imagens MODIS (do inglês, *Moderate Resolution Imaging Spectroradiometer*), selecionadas de uma área geográfica no leste do estado brasileiro do Mato Grosso, com cerca de 10 km^2 . O sistema desenvolvido foi uma rede neural artificial perceptron multicamadas, acrescida de modificações difusas e de um processo de filtragem de padrões para tratar o aspecto vago dos padrões analisados. Isso devido ao fato de a pesquisa não possuir padrões classificados suficientes para servir como conjunto de treinamento do sistema. Por isso, o primeiro passo deste trabalho foi o de classificar padrões, primeiramente por inspeção visual – tendo como base exemplos disponíveis na literatura – e, posteriormente, por um processo de filtragem dos padrões utilizando as médias e os desvios-padrão dos conjuntos de padrões classificados visualmente. A etapa de filtragem foi realizada com a finalidade de diminuir possíveis erros na classificação inicial, e se mostrou de grande utilidade, pois foi possível criar conjuntos de padrões separados em classes bem definidas, mesmo não existindo conhecimento especialista disponível. O segundo passo foi passar os padrões classificados e filtrados para quatro variações da rede perceptron multicamadas – com e sem adição de conceitos difusos –. No geral, todas as implementações obtiveram ótimos resultados, contudo, o uso das modificações difusas forneceu ao sistema resultados ainda melhores, diminuindo a influência de padrões erroneamente pré-classificados. Por fim, foi realizada a comparação entre a classificação obtida pelo sistema neuro-difuso e a classificação de uso do solo e cobertura da terra do projeto *TerraClass 2010*, o que permitiu verificar que o sistema desenvolvido obteve sucesso em seu objetivo, classificando corretamente a grande maioria dos novos padrões analisados.

Palavras-chave: Séries temporais. Índice de vegetação. Classificação de padrões. Uso do solo e cobertura da terra. Sistema neuro-difuso.

CLASSIFICATION OF TEMPORAL PATTERNS OF LAND USE AND LAND COVER IN TIME SERIES OF VEGETATION INDEX USING A NEURO-FUZZY SYSTEM

ABSTRACT

The question of changes in land use and land cover in Brazil have attracted great interest, both inside and outside the scientific community, mainly due to the potential environmental and socioeconomic impacts that these changes may cause. Pollution of air and water, desertification, illegal deforestation and biodiversity loss are just a few problems related to changes in land use and land cover. Therefore, all precautions should be taken for that local and global ecosystems are minimally affected. Based on this interest, it becomes very important to develop methods, mainly computational, able to monitor and identify changes in land use and land cover in the shortest possible time, since the processes of land degradation in areas of public interest are fast and, in the most cases, devastating. The remote sensing data have served as a crucial basis for this monitoring, however, usually the techniques for monitoring land use and land cover in Brazil require much time and expert knowledge to be performed, which limits the possibility of conservation actions by the competent authorities. That was the main motivation of this work, and the objective traced was to develop a system able of classifying temporal land use and land cover patterns using time series of vegetation index. We use time series obtained from MODIS images (Moderate Resolution Imaging Spectroradiometer), selected from a geographical area in eastern Brazilian state of Mato Grosso, with approximately 10 km^2 . The system developed is the multilayer perceptron neural network, plus diffuse modifications and a filtering process to treat the uncertainty aspect of the analyzed patterns. This is due to the fact that the study did not have enough classified patterns to serve as a training set. Therefore, the first step of this study was to classify patterns, first by visual inspection - based on examples from the literature - and, subsequently, by a filtering process of patterns using the means and standard deviations of the sets visually ranked. The filtering step was performed in order to reduce possible errors in the initial classification, and it proved to be very useful, because it was possible to create separated pattern sets into well-defined classes, even without available expert knowledge. The second step was to pass the filtered and classified patterns for four variations of multilayer perceptron network - with and without addition of fuzzy concepts -. In general, all implementations have had excellent results, however the use of diffuse changes provided even better results for the system, decreasing the influence of wrongly classified patterns. Finally, the comparison between the classification obtained by the neuro-fuzzy system and the land use and land cover classification of the *TerraClass 2010* project, has shown that the system was successful in its goal, correctly classifying the vast majority was held new standards analyzed.

Keywords: Time series. Vegetation index. Pattern classification. Land use and land cover. Neuro-fuzzy system.

LISTA DE FIGURAS

	<u>Pág.</u>
2.1 Ilustração de uma série temporal do índice de vegetação no pixel (i,j)	6
3.1 Esquema simplificado de um neurônio humano.	16
3.2 A estrutura geral de uma rede neural de múltiplas camadas.	17
3.3 Partição difusa do universo X representando “idade”, formada pelos conjuntos difusos “criança”, “adolescente” e “adulto”.	19
3.4 Funções de pertinência.	21
3.5 Duas regras SE-ENTÃO (R1 e R2), com duas funções de pertinência de variáveis de interesse, e métodos de inferência difusos comumente utilizados na literatura.	26
3.6 Sistema ANFIS.	28
4.1 Exemplo de grafo bipartido completo.	33
4.2 Exemplo de um espaço com duas classes sobrepostas.	39
5.1 Séries temporais de EVI2 com filtro de wavelet (vermelho) e sem filtro (azul) disponíveis no Laboratório Virtual de Sensoriamento Remoto < https://www.dsr.inpe.br/laf/series/ >; (a) conversão de floresta para pastagem em 2004; (b) conversão da floresta para outro tipo de vegetação com baixa biomassa em 2004 e em 2007-2008 introdução de agricultura; (c) mudança no sistema de cultivo da cana-de-açúcar com introdução de outra cultura em 2007-2008 e depois novamente cana-de-açúcar.	44
5.2 Intensificação no cultivo de arroz. As setas em vermelho indicam os picos de sazonalidade.	45
5.3 Representação gráfica das métricas de caracterização fenológica; no lado esquerdo datas chave para caracterização das sazonalidades. Na direita, métricas de caracterização da sazonalidade extraídas com a ferramenta TIMESAT.	46
5.4 Localização da área de estudo abordada.	47
5.5 Ilustração das observações desconsideradas.	48
6.1 Exemplo de degradação de um padrão temporal de uso do solo e cobertura da terra.	50
6.2 Comparativo preliminar entre métodos.	51
6.3 Classe Agricultura com a indicação dos valores de média + ou - $k = 1$ desvios-padrão ($k \times \sigma$).	52

6.4	Classes de séries temporais de índice de vegetação com indicação dos vetores média e um desvio padrão.	53
6.5	Duas das funções de ativação que podem ser usadas no MLP.	56
6.6	Resultados dos experimentos em 3 pixels, utilizando os conjuntos I (lado esquerdo) e II (lado direito).	63
6.7	Mapas de calor das taxas de acerto.	67
6.8	Amazônia legal e área de estudo.	68
6.9	Mapeamento TerraClass 2010.	68
6.10	Comparação entre classificações. a) Classificação TerraClass 2010 b) Classificação do Sistema Neural Difuso.	69

LISTA DE TABELAS

	<u>Pág.</u>
3.1 Principais $t - normas$ e $t - conormas$ duais.	23
3.2 Principais operadores de implicação.	24
6.1 Taxa de acerto individual para as classes F (floresta), D (desfloresta- mento), P (pastagem) e A (agricultura), usando os conjuntos I e II, e acerto global dos conjuntos I e II, para os métodos MLP tradicional (MLP), MLP utilizando grau de ambiguidade (MLP-G), MLP usando saída difusa desejada(MLP-D) e MLP utilizando ambos, grau de ambi- guidade e saída difusa desejada (MLP-GD).	61
6.2 Percentual de vezes que cada quantidade de neurônios na camada escondida obteve as melhores taxas de acerto.	64

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO	1
1.1 Objetivos e resultados esperados	3
1.2 Organização da dissertação	4
2 SÉRIES TEMPORAIS EM SENSORIAMENTO REMOTO . .	5
2.1 Séries temporais de índice de vegetação	5
2.2 Análise de séries temporais	7
2.2.1 Algoritmos de reconhecimento de padrões	9
2.2.1.1 Métodos com aprendizagem supervisionada	9
2.2.1.2 Aprendizagem semi-supervisionada	11
2.2.1.3 Aprendizagem não supervisionada	11
2.2.2 Método de classificação escolhido	13
3 SISTEMAS NEURO-DIFUSOS	15
3.1 Redes neurais artificiais	15
3.2 Sistemas difusos	18
3.2.1 Conjuntos clássicos	18
3.2.2 Conceitos básicos da teoria dos conjuntos difusos	20
3.2.2.1 Operadores de conjuntos difusos	22
3.2.2.2 Modelos de inferência difusos	24
3.3 Sistemas neuro-difusos	26
4 PERCEPTRON MULTICAMADAS DIFUSO	31
4.1 Introdução	31
4.2 A rede perceptron multicamadas	32
4.2.0.3 Função de ativação	33
4.2.0.4 O processo de treinamento da rede	34
4.3 O perceptron multicamadas difuso	36
4.3.1 Fuzzificação da rede MLP	37
4.3.1.1 Vetor de saída desejada	38
4.3.1.1.1 - Cálculo da saída desejada difusa	39
4.3.1.2 O processo de atualização difusa dos pesos	40

5 IDENTIFICAÇÃO DE PADRÕES TEMPORAIS DE USO E COBERTURA DA TERRA	43
5.1 Classificação inicial dos dados	46
6 MÉTODOS E RESULTADOS	49
6.1 Testes preliminares	49
6.2 Filtragem dos dados	51
6.3 Aspectos práticos do treinamento de redes MLP	54
6.4 Definição dos parâmetros do sistema	60
6.5 Comparação entre métodos e entre conjuntos de dados	60
6.6 Estudo paramétrico	63
6.7 Comparação com classificações TerraClass 2010	65
7 CONCLUSÕES	71
REFERÊNCIAS BIBLIOGRÁFICAS	73

1 INTRODUÇÃO

Os conceitos de *uso do solo* e *cobertura da terra* são similares, mas não equivalentes. O termo *cobertura da terra* diz respeito à caracterização do estado físico, químico e biológico da superfície terrestre, como floresta, água e área construída. Já, *uso do solo*, compreende os propósitos humanos associados àquela cobertura. No caso da cobertura “floresta”, por exemplo, dois tipos de uso podem ser: área de conservação e área de recreação (BRIASSOULIS, 2000). Uma única classe de cobertura pode suportar múltiplos usos, ao mesmo tempo que um único sistema de uso pode incluir diversas coberturas.

As mudanças no uso do solo e cobertura da terra têm despertado atenção crescente em todo o mundo, dentro e fora do meio científico. Isso se deve ao acelerado processo de mudanças das últimas décadas, assim como aos possíveis impactos ambientais e socioeconômicos que estas mudanças podem acarretar, que causam preocupações desde o nível local até o global do meio ambiente e da sociedade.

De forma global, são questões de interesse: o relacionamento entre os padrões de uso do solo e cobertura da terra e o aquecimento global, o aumento do buraco na camada de ozônio e o aumento do nível do mar. Em termos locais, as questões ambientais relacionadas a mudanças no uso do solo e cobertura da terra são as conhecidas: poluição do ar e da água, degradação do solo, desertificação, perda de biodiversidade, entre outras. Além do interesse ambiental, questões relacionadas a disponibilidade de alimentos e de água para a crescente população mundial, as migrações humanas, e as questões de segurança frente a alterações causadas por fenômenos naturais ou mudanças tecnológicas são temas de alta relevância na nossa sociedade (AGUIAR, 2012).

É importante dizer que, embora sejam principalmente os impactos negativos que motivam o interesse pelo entendimento dos processos de mudança no uso e cobertura do solo, nem sempre as mudanças são negativas, especialmente se consideradas a escala temporal e espacial de observação do evento e suas consequências (BRIASSOULIS, 2000). Segundo Aguiar (2012), a questão mais importante que se coloca atualmente a esse respeito é a sustentabilidade do desenvolvimento humano e tecnológico, e o balanço adequado entre as questões sociais, econômicas e ambientais envolvidas.

Devido a essas preocupações, se torna de grande importância o desenvolvimento e utilização de ferramentas e mecanismos capazes de identificar classes de uso do solo e cobertura da terra, tanto em escala espacial – considerando grandes áreas

de interesse –, quanto em escala temporal – considerando seu histórico no decorrer do tempo. Isso para entender os processos de mudanças de uso do solo e cobertura da terra e seus principais fatores determinantes, assim como suas consequências socioeconômicas e ambientais.

Uma das áreas de estudo que mais vem disponibilizando dados para um monitoramento preciso sobre a dinâmica da ocupação do solo é o *Sensoriamento Remoto*, que, em resumo, é a tecnologia que permite obter imagens – e outros tipos de dados – da superfície terrestre, por meio da captação e registro da energia refletida ou emitida pela superfície. Na área de sensoriamento remoto, os estudos de mudanças de uso e cobertura da terra em escalas local ou global, utilizando imagens multitemporais derivadas de sensores a bordo dos satélites de observação da Terra, têm despertado grande interesse na comunidade científica, pois têm permitido não só a visualização, mas a análise das causas e efeitos decorrentes das mudanças de uso do solo e cobertura da terra (DEFRIES, 2008).

Para as análises de mudanças no uso do solo e cobertura da terra, os índices de vegetação derivados de sensores orbitais são comumente utilizados, pois apresentam informações de fácil interpretação, e são sensíveis a alterações na cobertura vegetal. O pressuposto básico para utilização desses índices é de que se há alteração no regime dinâmico de variabilidade da vegetação, então possivelmente existe alguma mudança de origem antrópica no uso solo e cobertura da terra. Um exemplo clássico é a supressão da vegetação num processo de desflorestamento, em que os valores de índice de vegetação diminuem abruptamente quando não existe nenhum tipo de vegetação no solo, o chamado corte raso. Mais de uma dezena de índices de vegetação são utilizados para ressaltar os contrastes entre tipos e aspectos fenológicos da vegetação, dependendo do tipo de aplicação. Em geral, a utilização desses índices requer conhecimento adicional em sensoriamento remoto para a seleção e uso de bandas espectrais adequadas para o seu cálculo (SKOLE; TUCKER, 1993; DEFRIES et al., 2002; DEFRIES, 2008).

Os índices de vegetação extraídos a partir de imagens de satélite, tomadas de uma mesma localização geográfica, podem ser agrupados de modo a se criar séries temporais de índice de vegetação. Muitos trabalhos utilizam dessas séries, como Barbosa (1998), Wardlow et al. (2007) e Lacruz e Júnior (2011). Entretanto, em geral, cada trabalho sintetiza suas próprias séries temporais a partir das imagens de satélite. Freitas (2012), por outro lado, foi responsável por um grande avanço nesse aspecto, pois desenvolveu uma ferramenta de laboratório virtual de sensoriamento remoto

que permitiu aos usuários terem acesso às séries temporais de índice de vegetação de todo o território do continente sul-americano, economizando muito tempo e trabalho para outros pesquisadores.

As séries temporais de índice de vegetação possuem um grande valor agregado, pois, a partir delas, é possível se estudar as variações do verde na superfície terrestre, num longo período de tempo, e relacionar quais fenômenos produzem tais variações, ou seja, é possível classificar mudanças de uso do solo e cobertura da terra a partir de séries temporais de índice de vegetação. Essa classificação por um especialista humano é possível, porém árdua e demorada quando existe uma grande quantidade de dados, por isso a criação de um sistema computacional, que seja capaz de realizar essa tarefa instantaneamente ao se apresentar novos casos, se torna essencial para o monitoramento eficaz da superfície terrestre. Além disso, são muitos os interessados por um sistema como esse, tais como gerentes e empresários dos ramos de agricultura e pecuária, pesquisadores científicos, agentes ambientais e também agricultores familiares.

1.1 Objetivos e resultados esperados

Com base nesse contexto, o objetivo principal deste trabalho é o de criar um sistema computacional que seja capaz de classificar padrões temporais de uso do solo e cobertura da terra em séries temporais de índice de vegetação. Para isso, é preciso cumprir os seguintes objetivos secundários:

- Criar um conjunto classificado de padrões temporais de uso do solo e cobertura da terra para servir como base de conhecimento;
- Filtrar o conjunto inicialmente criado a fim de eliminar possíveis ruídos;
- Implementar um sistema computacional de classificação de padrões de uso do solo e cobertura da terra;
- Criar um índice de confiança para mostrar o quão certo o sistema está na classificação de novos casos.

Como resultados, espera-se que consigamos criar uma base de conhecimento confiável, implementar o sistema de classificação e que este sistema classifique, de maneira correta, a grande maioria dos padrões apresentados. Além disso, o sistema deve indicar a quantidade de confiança que possui para as classificações apresentadas.

Os passos definidos como necessários para cumprir os objetivos propostos são:

- Estudo sobre o problema e sobre as áreas que o compõem;
- Definição de qual sistema computacional de classificação será utilizado para a tarefa proposta;
- Obtenção das séries temporais de índice de vegetação;
- Divisão das séries temporais em padrões de uso do solo e cobertura da terra;
- Classificação visual dos padrões obtidos;
- Filtragem dos padrões inicialmente classificados;
- Implementação do sistema de classificação escolhido;
- Desenvolvimento de uma métrica para determinar o quão certo o sistema está da classificação apresentada para novos padrões;
- Realização de testes para verificar a qualidade do sistema;
- Comparação dos resultados apresentados pelo sistema proposto com a classificação de outra fonte.

1.2 Organização da dissertação

Esta dissertação organiza-se em 7 capítulos. O presente capítulo trata da introdução ao problema e da solução proposta. No Capítulo 2 são apresentados os conceitos sobre séries temporais de índice de vegetação e métodos de análise de séries temporais potencialmente capazes de resolver o problema de classificar padrões temporais de índice de vegetação. Já, no Capítulo 3, são apresentados os sistemas neuro-difusos, base do sistema de classificação implementado para o objetivo proposto. No Capítulo 4, mostra-se a fundamentação teórica sobre o sistema de classificação desenvolvido. Em seguida, o Capítulo 5 mostra como o conjunto de padrões classificados – necessários para o treinamento – foi inicialmente gerado. O Capítulo 6 possui a descrição de como as técnicas e métodos propostos foram utilizados. Além disso, traz os resultados obtidos neste trabalho e as análises referentes a esses resultados. Finalmente, no Capítulo 7 são colocadas algumas considerações e desafios em aberto na área de monitoramento e detecção de mudança de uso do solo e cobertura da terra por meio de séries temporais de alta resolução.

2 SÉRIES TEMPORAIS EM SENSORIAMENTO REMOTO

Ao longo dos séculos, as pessoas têm se fascinado por compreender os acontecimentos que variam com o tempo. Registros da vazão anual do rio Nilo, por exemplo, existem desde o ano 622 depois de Cristo (PUTTER *et al.*, 1998), e fenômenos astrofísicos, como o número de manchas solares, foram registradas desde o século XVII. Em geral, os fenômenos de interesse público são observados ao se olhar suas principais variáveis ao longo do tempo - de forma contínua ou discreta. As ondas cerebrais, por exemplo, são representadas como eventos ordenados no tempo e eletrocardiogramas produzem traços, baseados no tempo, de ondas do coração. Em meteorologia, a velocidade do vento, temperatura, pressão, umidade e medições de precipitação ao longo do tempo estão associadas às condições meteorológicas. Dados da produção industrial, taxas de juros, inflação, preços de ações e taxas de desemprego, entre outros dados colhidos no decorrer do tempo, podem fornecer uma medida da saúde de uma economia.

Esse conjunto de observações, feitas sequencialmente ao longo do tempo, é denominado série temporal. Como dito anteriormente, as séries temporais são utilizadas para caracterizar a evolução no tempo do comportamento de uma grande variedade de sistemas. Segundo Lin *et al.* (2012), uma característica intrínseca de uma série temporal é que, geralmente, as observações adjacentes são interdependentes. A natureza dessa interdependência é de grande interesse no campo de análise de séries temporais, que se preocupa com as técnicas e modelos dinâmicos sobre a natureza dos dados das séries e o uso de tais modelos nas mais diversas áreas de aplicação (LIN *et al.*, 2012).

2.1 Séries temporais de índice de vegetação

As séries temporais de dados relacionados ao sensoriamento remoto permitem detectar e monitorar mudanças na superfície da terra. Um exemplo disso é a existência de aplicações – mesmo com dados de baixa resolução temporal – que avaliam variações inter-anuais da cobertura terrestre. Nesse contexto está o projeto PRODES (Programa de Cálculo do Desflorestamento da Amazônia) – desenvolvido pelo Instituto Nacional de Pesquisas Espaciais (INPE) –, que produz estimativas das taxas anuais de desflorestamento na Amazônia Legal a partir de dados de satélite (CÂMARA *et al.*, 2006). Numa escala de tempo maior, estudos sobre a dinâmica inter e intra-anual dos ecossistemas terrestres demandam dados de sensoriamento remoto com alta frequência temporal. Esses sistemas refletem as variações sazonais de variáveis climáticas, incluindo temperatura e precipitação, e respondem também às mudanças climáticas de longo prazo (ZHANG *et al.*, 2003).

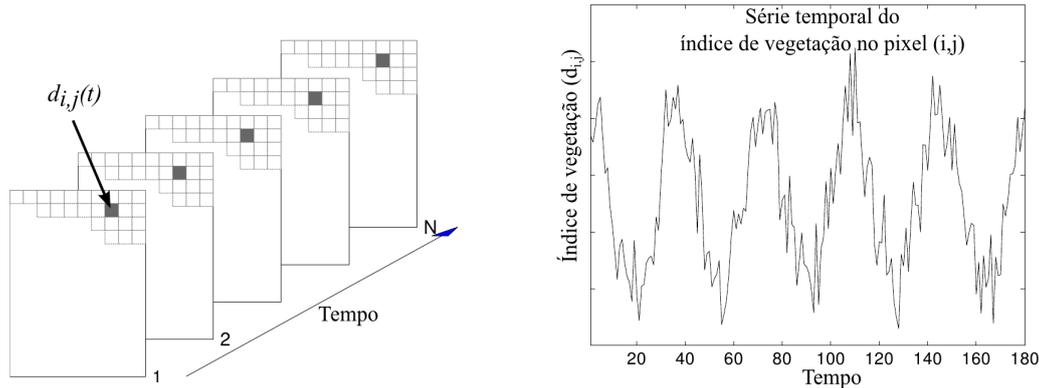


Figura 2.1 - Ilustração de uma série temporal do índice de vegetação no pixel (i,j).

Fonte: Maus (2013).

Nas análises de dinâmica de vegetação, normalmente, calculam-se índices de vegetação – que são transformações lineares de bandas espectrais, geralmente nas faixas do vermelho (V) e infravermelho próximo (IVP) do espectro eletromagnético (EPIPHANIO et al.,) – baseados em dados de satélites, que além de realçar as variações na biomassa verde, ajudam a reduzir interferências indesejadas (TUCKER, 1979). Os índices de vegetação calculados e extraídos de imagens de satélite com boa e confiável resolução temporal podem ser agrupados e analisados de modo a se criar séries temporais de índice de vegetação. A Figura 2.1 ilustra o emprego desses índices na composição de séries temporais. Observa-se, para cada pixel da área em questão, uma série temporal que caracteriza a variação da vegetação ao longo do tempo. Para que uma composição temporal desse tipo seja consistente é essencial que os dados multi-temporais obtidos por satélite estejam georreferenciados.

Tradicionalmente, existem diversos estudos que usam imagens de satélite tomadas de uma mesma área em diferentes épocas para detecção de mudanças no estado de um objeto ou área (HAYES; SADER, 2001). Diferentes fenômenos podem ser identificados por meio da detecção de mudanças, como desflorestamentos, modificações no uso do solo (substituição de matas nativas por agricultura), queimadas, variações na geometria de aquisição das imagens, etc. Estas mudanças resultam em alterações nos valores de radiância dos pixels – o que afeta o índice de vegetação – (SILVA et al., 2012), possibilitando a comparação de duas datas ou, no caso de séries temporais de índice de vegetação, de vários conjuntos de datas.

A maior frequência temporal de amostragem produz dados de sensoriamento remoto

com potencial para analisar características sazonais da vegetação, ou seja, permite analisar suas flutuações intra-anuais. Dados com essa característica viabilizam o monitoramento e mapeamento da cobertura vegetal com base em características fenológicas (LLOYD, 1990). Dessa forma, longas séries temporais de dados com alta resolução temporal podem contribuir para a compreensão da relação clima-biosfera, já que a dinâmica fenológica dos ecossistemas terrestres reflete a dinâmica intra e inter-anual do clima e do regime hidrológico do planeta (MYNENI et al., 1997; ZHANG et al., 2003).

Até o final da década de 1990 o sensor AVHRR foi a única fonte de dados globais com longas séries temporais disponíveis para estudo da vegetação (JAMES; KALLURI, 1994; ZHANG et al., 2003). Entretanto, devido ao fato de a finalidade principal do lançamento desse sensor não ser o estudo da dinâmica global de vegetação, seus dados não são muito adequados para o monitoramento de vegetação (GOWARD et al., 1991). Por outro lado, o sensor MODIS (*Moderate Resolution Imaging Spectroradiometer*), planejado para monitoramento do oceano, atmosfera e também da terra, disponibiliza, desde 2000, índices de vegetação desenvolvidos para um preciso monitoramento sazonal e inter-anual da vegetação (JUSTICE et al., 1998).

Os produtos gerados por meio do sensor MODIS permitiu avanços nos trabalhos que necessitam de maior resolução temporal, quase diária, aliada à moderada resolução espacial (250 m). Por esse motivo, e pela qualidade do georreferenciamento das imagens MODIS, é possível elaborar séries temporais com precisão (WOLFE et al., 2002), que se tornam eficazes ferramentas para identificar processos de conversão do uso do solo e cobertura da terra, como visto nos trabalhos Nassar et al. (2008), Rudorff et al. (2009) e Rudorff et al. (2010), que avaliam mudanças de uso do solo e cobertura da terra em decorrência da expansão do cultivo da cana-de-açúcar na região Centro-Sul do Brasil.

2.2 Análise de séries temporais

Uma importante área de aplicação do uso de séries temporais é a previsão de valores futuros. Por exemplo, o uso, no tempo t , de observações disponíveis de uma série temporal para prever o seu valor, em algum tempo futuro $t + l$, pode fornecer uma base para o planejamento econômico e de negócios, planejamento de produção, inventário e controle da produção, controle e otimização de processos industriais, entre outros.

Outro tema de interesse considerável em análise de séries temporais é a estimativa

de funções de transferência, que busca, utilizando modelos matemáticos, determinar explicitamente como as observações de uma série relacionam-se com seus valores passados. No geral, esses modelos matemáticos ajudam na inferência e interpretação das quatro componentes que constituem o modelo clássico de séries temporais:

- componente tendência (T) – é o comportamento de longo prazo da série –,
- componente cíclica (C) – são as flutuações cíclicas nos valores da variável de média a longa duração –,
- componente sazonal (S) – são as flutuações, de curta duração, nos valores da variável de interesse – e
- componente aleatória (A) – são as variações aleatórias da variável.

As quatro componentes citadas podem definir a equação da série temporal, para a variável de interesse Y , como $Y = f(T, C, S, A)$ (CHATFIELD, 2013).

Um problema usual em séries temporais ocorre em aplicações com grandes bancos de dados. Com isso, há o crescimento do interesse no desenvolvimento de algoritmos e técnicas de mineração de dados e reconhecimento de padrões. Segundo Lin et al. (2012), as características estruturais intrínsecas dos dados de séries temporais, como a alta dimensionalidade e a correlação característica, combinadas com os ruídos induzidos pelas medições de séries temporais do mundo real, são os desafios que tornam os algoritmos clássicos de mineração de dados, frequentemente, ineficazes e ineficientes para busca de padrões.

Em mineração de dados, técnicas de *clustering*, ou agrupamento de padrões, são utilizadas para buscar grupos de padrões em conjuntos de dados de séries temporais, por meio de alguma medida de similaridade. Trabalhos como os de Aach e Church (2001), Kalpakis et al. (2001) e Keogh e Pazzani (1998) podem ser citados como exemplos de estudos de agrupamento. Já, trabalhos como os de Santana et al. (2010), Adami et al. (2008), Formigoni et al. (2011) e Lacruz et al. (2005) tratam de uma área similar, que é a classificação de padrões em séries temporais de índice de vegetação. A classificação difere do agrupamento de padrões, principalmente, no sentido que o primeiro deve possuir informação a priori sobre o comportamento das observações das séries, ou seja, deve existir conhecimento especialista sobre como a série evolui para determinados cenários. Em contraponto, é possível determinar qual o cenário existente para certo subconjunto de observações, de acordo com o comportamento dos dados desse subconjunto.

Na classificação de padrões, tem-se um conjunto de padrões que foram anteriormente classificados, denominado *conjunto de treinamento*, e, com um mecanismo de extração de características, que computa informações numéricas ou simbólicas, é possível classificar novos padrões com a comparação do conteúdo desse padrão com o conjunto de características extraídas do treinamento.

Neste trabalho, estudamos algumas abordagens de classificação de padrões para o objetivo proposto de classificar padrões temporais de uso do solo e cobertura da terra em séries temporais de índice de vegetação. Apresentamos essas abordagens na Seção 2.2.1.

2.2.1 Algoritmos de reconhecimento de padrões

O reconhecimento de padrões é o processo de mapeamento automático da representação de entrada de uma entidade a uma dada classe. A tarefa de reconhecimento é geralmente categorizada com base em como o processo de aprendizagem determina a classe de saída. Esse procedimento de aprendizagem pode ser supervisionado – quando um determinado padrão é atribuído a uma das classes pré-definidas do problema, usando dados rotulados para construir um modelo ou para orientar a classificação de padrões –, não supervisionada – quando um determinado padrão é atribuído a uma classe desconhecida –, ou semi-supervisionada – quando um determinado padrão é atribuído a uma das classes pré-definidas, usando tanto dados rotulados quanto dados não rotulados. Para a aprendizagem supervisionada, como classificação, um modelo funcional é muitas vezes utilizado para mapear as entradas observadas às classes de saída. Uma grande quantidade de técnicas de construção de modelos têm sido desenvolvidas para esta finalidade (DUDA et al., 2012), incluindo: árvores de decisão, indução de regras, redes bayesianas, raciocínio baseado em memória, máquinas de vetores de suporte e redes neurais artificiais.

2.2.1.1 Métodos com aprendizagem supervisionada

Métodos supervisionados, como dos k Vizinhos Mais Próximos (k -NN, do inglês *k Nearest Neighbor*), têm sido utilizados com sucesso para a classificação de dados de séries temporais (LIN et al., 2012). O algoritmo de classificação NN trabalha com o cálculo da distância entre o objeto a ser classificado e cada membro do conjunto de treinamento (HAN et al., 2006). A classificação do objeto de interesse é prevista para ser a mesma que a classificação dos k membros do conjunto de treinamento mais próximo. O classificador 1-NN, com validação cruzada *leaving-one-out* (deixando um para fora), tornou-se o método padrão usado para comparar e avaliar a utilidade de

representações de séries temporais e medidas de similaridade (DING et al., 2008).

Outra opção comum de classificador é a árvore de decisão (HAN et al., 2006). Trata-se de uma ferramenta clássica de aprendizado de máquina que usa uma estrutura de árvore parecida com um fluxograma, em que um nó interno denota um teste em um atributo, um ramo denota o resultado em um teste, e um nó folha denota a classe rótulo ou distribuição de classes. Ao colocar o atributo que melhor distingue os dados na raiz – ou seja, o que resulta em maior ganho de informação –, um processo de indução de árvore de decisão particiona recursivamente as amostras de dados em subconjuntos de amostras de acordo com algum critério de divisão. A árvore resultante pode ser utilizada para classificar as futuras amostras de entrada, e também pode ser facilmente convertida em um conjunto de regras que generalizam o comportamento dos dados. Mesmo sendo uma técnica utilizada com sucesso em classificação de padrões, as árvores de decisão enfrentam algumas dificuldades tratando de dados de séries temporais, uma vez que o elevado nível de ruído e dimensionalidade pode resultar numa árvore profunda, espessa e com pouca precisão (LIN et al., 2012). Na tentativa de superar esse problema, Geurts (2001) sugeriu representar a série temporal como uma árvore de regressão (RT, do inglês *Regression Tree*), e treinar a árvore de decisão diretamente sobre essa representação.

As Redes Neurais Artificiais (RNAs) (LIPPMANN, 1987) são modelos inspirados na estrutura e no funcionamento do sistema nervoso e são úteis na solução de diversos problemas computacionais, dentre eles, o reconhecimento de padrões. Segundo Haykin (2001), uma rede neural é um processador maciçamente paralelamente distribuído constituído de unidades de processamento simples – os neurônios –, e que tem a propensão natural para armazenar conhecimento experimental e torná-lo disponível para o uso. A RNA se assemelha ao cérebro humano pelo fato de o conhecimento ser adquirido pela rede, a partir de seu ambiente, através de um processo de aprendizagem; e também por serem utilizadas forças de conexão entre os neurônios, conhecidas como pesos sinápticos, para se armazenar o conhecimento adquirido. Durante o treinamento, cada exemplo é alimentado através da rede, e os pesos são atualizados de acordo com a diferença entre a saída real e a saída esperada. As RNAs têm sido aplicadas com sucesso em domínios tais como reconhecimento de caracteres (DESAI, 2010), detecção de faces (AITKENHEAD; MCDONALD, 2003), reconhecimento de voz (DEDE; HÜSNÜ, 2010) e astronomia (COLLISTER; LAHAV, 2004; WANG et al., 2009).

As Máquinas de Vetores de Suporte (SVMs, do inglês *Support Vector Machines*) (DUDA et al., 2012) são comumente usadas para a construção de modelos de classi-

ficação para representações de alta dimensionalidade. As observações do conjunto de treinamento que melhor definem a fronteira de decisão entre as classes são escolhidas como vetores de suporte, que suportam/definem o limite de decisão. Um peso diferente de zero é atribuído a cada vetor de suporte. A classificação de um objeto de teste será realizada calculando uma soma ponderada das saídas da função de similaridade entre o novo objeto e os vetores de suporte de cada classe, sendo que a classe predita será a classe com a maior soma. As SVMs têm sido utilizadas por vários pesquisadores em classificação de séries temporais. [Kampouraki et al. \(2009\)](#) extraíram características a partir de séries temporais utilizando métodos estatísticos e técnicas de análise de sinais, e classificaram os dados usando SVMs. Já o trabalho de [Eads et al. \(2002\)](#) propôs o uso de algoritmos genéticos para extrair características de séries temporais e classificá-las usando SVMs.

2.2.1.2 Aprendizagem semi-supervisionada

Todas as técnicas descritas até aqui são técnicas de aprendizagem supervisionada, onde padrões com rótulos de classificação conhecidos são usados para construir um modelo para identificar padrões com rótulos de classificação desconhecidos. A aprendizagem semi-supervisionada, por outro lado, envolve o uso, tanto de padrões rotulados, quanto de padrões não rotulados. Para séries temporais, [Wei e Keogh \(2006\)](#) propuseram uma técnica de classificação semi-supervisionada baseada no 1-NN. O algoritmo começa por treinar o classificador usando todos os padrões rotulados. Em seguida, ele classifica os dados não rotulados, e adiciona no conjunto de treinamento os dados não rotulados com classificação mais confiável ([WEI; KEOGH, 2006](#)). O processo é repetido até que algum critério de parada seja atingido. Os autores demonstraram que seu algoritmo requer apenas uma pequena quantidade de dados rotulados para construir um classificador preciso.

2.2.1.3 Aprendizagem não supervisionada

Ao contrário da aprendizagem supervisionada e da aprendizagem semi-supervisionada, a aprendizagem não supervisionada (por exemplo, o agrupamento de padrões) envolve apenas dados não rotulados. Uma das abordagens do agrupamento de padrões mais utilizadas é o agrupamento hierárquico, devido ao grande poder de visualização que oferece ([KEOGH; PAZZANI, 1998](#)). O agrupamento hierárquico produz uma hierarquia aninhada de grupos de objetos semelhantes, de acordo com uma matriz de distâncias dos objetos. Uma das vantagens desse método é a sua generalidade, uma vez que o usuário não precisa fornecer todos os parâmetros do sistema, tal como o número de grupos. No entanto, devido à sua complexidade

computacional, no mínimo quadrática, a sua aplicação fica limitada a conjuntos de dados pequenos ou ao uso de computadores com alta capacidade de processamento e memória (LIN et al., 2012).

Um método mais rápido e simples para realizar o agrupamento de padrões é o k-médias (MACQUEEN, 1967; BRADLEY et al., 1998). A intuição básica por trás do k-médias – e, em geral, dos algoritmos de refinamento iterativo – é a mudança contínua dos objetos nos grupos, de modo que a distância dentro do grupo seja minimizada. O algoritmo do k-médias para N objetos tem uma complexidade $O(kNrD)$ (MACQUEEN, 1967), onde k é o número de grupos especificados pelo usuário, r é o número de iterações até a convergência, e D é o número de atributos usados. Pelo fato do k-médias ser essencialmente um algoritmo do tipo *hill-climbing*, é garantido que ele convirja para um ótimo local, mas não necessariamente para um ótimo global. Com isso, as escolhas dos centros iniciais são críticas para a qualidade dos resultados (LIN et al., 2012).

Outro algoritmo bem conhecido de agrupamento divisivo é o algoritmo de Maximização de Expectativa (EM, do inglês *Expectation-Maximization*). O algoritmo EM é muito semelhante ao k-médias, visto que também começa com uma estimativa inicial quanto à localização dos grupos. A principal diferença entre estes algoritmos é que o k-médias tenta modelar os dados como uma coleção de k regiões esféricas, com cada objeto pertencendo a exatamente um grupo. Em contraste, o EM modela os dados como uma coleção de k distribuições, normalmente gaussianas, tendo cada um dos objetos de dados um grau de pertinência para cada grupo (LIN et al., 2012). A principal vantagem do EM sobre o k-médias é a sua capacidade para modelar um conjunto muito mais rico de grupos (DEMPSTER et al., 1977; LAWRENCE; REILLY, 1990).

Algoritmos de agrupamento especializados têm sido propostos para utilização em análises de séries temporais. Por exemplo, o trabalho de Lin et al. (2004) propôs uma versão incremental e iterativa do algoritmo k-médias chamado i-kMédias. O algoritmo funciona aproveitando a propriedade de multi-resolução de *wavelets*, que elimina o dilema de seleção de centros iniciais do k-médias (LIN et al., 2004). Rodrigues et al. (2008) propôs um algoritmo de agrupamento hierárquico para os fluxos de dados de séries temporais. O algoritmo incremental constrói uma hierarquia de árvore de grupos, usando uma medida de dissimilaridade baseada na correlação entre as séries temporais (RODRIGUES et al., 2008).

2.2.2 Método de classificação escolhido

Como visto na seção anterior, existem diversos métodos teoricamente capazes de resolver o problema de classificar padrões temporais de uso do solo e cobertura da terra. O primeiro limitante é que em algum momento da aplicação do método escolhido seria necessária a utilização de conhecimento especialista. Assim, um método que utiliza, no mínimo, a aprendizagem semi-supervisionada seria o necessário. Entre as opções disponíveis, consideramos que uma técnica que utilizasse de aprendizagem supervisionada seria não só capaz, mas preferida, devido, principalmente, pelo desafio de tratar dados inicialmente não classificados por um especialista, nos deixando, dessa forma, com dados incertos. O desafio de lidar com dados incertos vem do fato de os exemplos disponíveis na literatura serem escassos, mas suficientes para nos orientar sobre como os padrões temporais de índice de vegetação – de algumas classes – evoluem com o tempo.

Foi escolhida aqui uma técnica que mistura a natureza conexionista e a aprendizagem supervisionada da rede neural artificial Perceptron Multicamadas, com a natureza do tratamento da incerteza da lógica difusa. A intenção foi a de utilizar os exemplos disponíveis na literatura para comparar com os padrões presentes na área de estudo e separar esses padrões em classes pré-definidas. Essa separação foi realizada por um não especialista e a existência de ruídos nos conjuntos de padrões classificados motivou a utilização da lógica difusa juntamente com a rede neural.

3 SISTEMAS NEURO-DIFUSOS

No campo da inteligência artificial, neuro-difuso refere-se à combinação de redes neurais artificiais e lógica difusa. Ele resulta em um sistema híbrido inteligente que combina o estilo de raciocínio humano de sistemas difusos com a estrutura conexionista e de aprendizagem de redes neurais, sendo uma solução promissora para diversos problemas do mundo real e já um sucesso em áreas como reconhecimento de padrões e controle (CANUTO, 2001). Esses sistemas incorporam o estilo de raciocínio humano através da utilização de um quadro estrutural de alto nível, de pensamento e raciocínio, com o uso de regras difusas SE-ENTÃO (JIN, 2000). A rigor, qualquer sistema que misture os paradigmas de sistemas difusos e sistemas conexionistas poderia ser chamado de neuro-difuso (CANUTO, 2001).

Neste capítulo, apresenta-se os conceitos básicos de sistemas neuro-difusos, assim como das áreas do conhecimento que as compõem: as Redes Neurais Artificiais (RNAs) e a Lógica Difusa.

3.1 Redes neurais artificiais

De acordo com Lent (2004), o cérebro humano consiste de aproximadamente 10^{11} neurônios (unidades de processamento), com cerca de 10^{14} conexões de um neurônio para o outro. A Figura 3.1 mostra o esquema simplificado de um neurônio humano. A célula em si é constituída por um núcleo e o exterior é uma membrana celular. Cada neurônio tem um nível de ativação, que varia entre um máximo e um mínimo. As conexões (sinapses) existem para aumentar (excitando) ou diminuir (inibindo) a ativação através de outros neurônios. Essas sinapses transmitem o nível de ativação de um neurônio para outro (MÜLLER et al., 2013).

Apesar da baixa operação de neurônios individuais, o cérebro pode lidar eficientemente com tarefas computacionais complexas. Além disso, dados incompletos e inconsistentes podem ser manipulados pelo cérebro, que pode aprender com a experiência e é tolerante a falhas.

O tema principal da investigação de redes neurais artificiais se concentra na modelagem do cérebro como um dispositivo de computação paralela para várias tarefas computacionais que são difíceis de resolver usando métodos computacionais convencionais (MEHROTRA et al., 1996).

As redes neurais artificiais (HAYKIN, 2001) são modelos inspirados na estrutura e no funcionamento do sistema nervoso e são úteis na solução de diversos problemas

computacionais, dentre eles: reconhecimento de padrões, comportamento adaptativo, probabilísticos, fusão de dados, entre outros. Além disso, as redes neurais são uma promessa de processamento eficiente, explorando paralelismo maciço e comportamento tolerante a falhas.

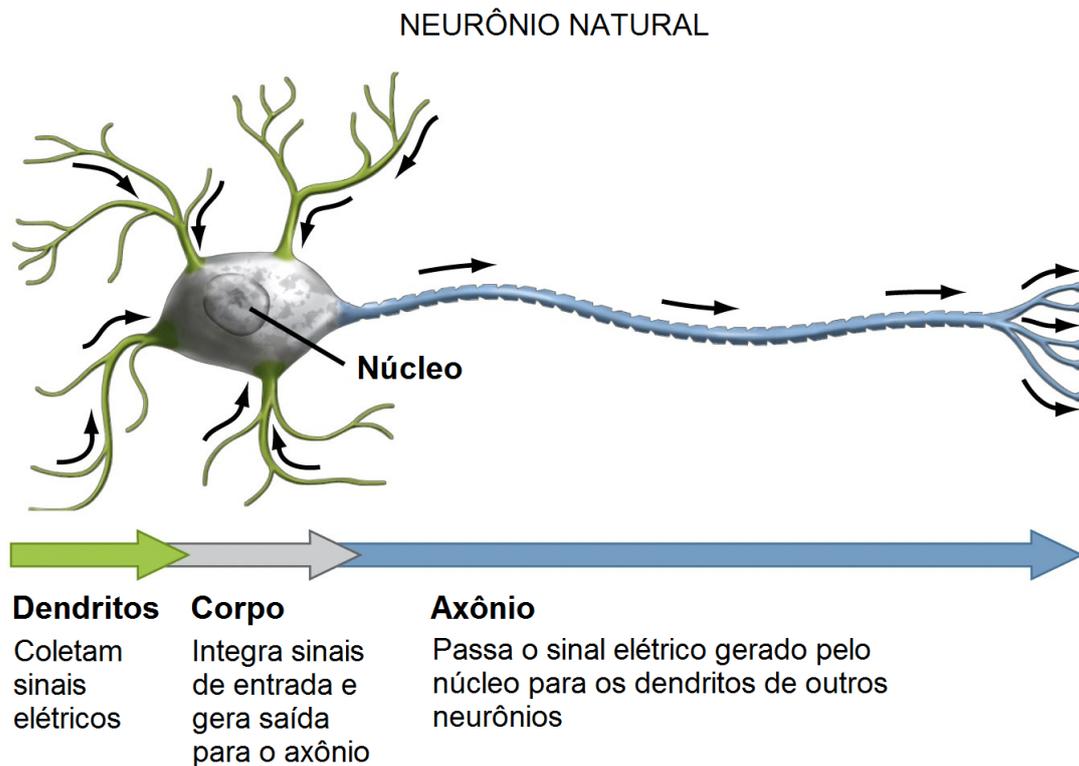


Figura 3.1 - Esquema simplificado de um neurônio humano.

Fonte: Adaptado de Freeman (2005).

Elas foram motivadas, desde sua concepção, a simular o tipo de processamento que acontece no cérebro humano, que processa informações de uma forma totalmente diferente do computador digital convencional, sendo considerado um computador extremamente complexo, não linear e paralelo, com capacidade superior a de qualquer computador atual.

O processo de aprendizagem, também chamado de algoritmo de aprendizagem, tem o objetivo de alterar os pesos das conexões entre neurônios de forma a alcançar a máxima aproximação possível da saída desejada. No caso da rede neural Perceptron

Multicamadas (MLP, do inglês *Multilayer Perceptron*), especificamente, tal objetivo é minimizar o erro quadrático relacionado à diferença entre os resultados obtido e esperado (HAYKIN, 2001).

A Figura 3.2 mostra a estrutura, em forma de grafo, de uma rede neural artificial composta por quatro camadas: camada de entrada, camada de saída e duas camadas intermediárias, também conhecidas como camadas escondidas. Os nós do grafo são os neurônios e as arestas que ligam os nós são as conexões da rede.

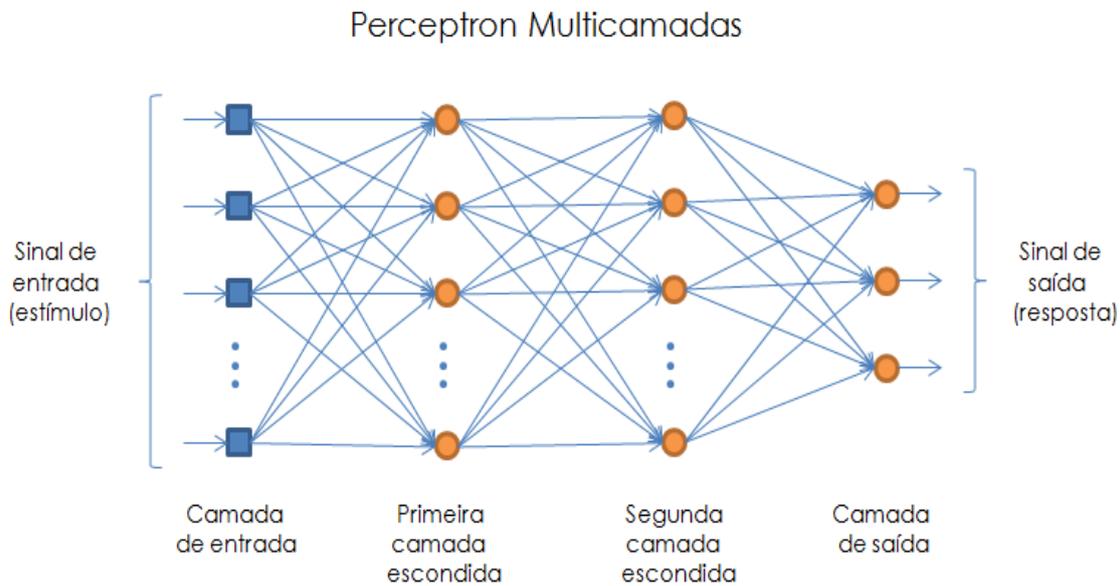


Figura 3.2 - A estrutura geral de uma rede neural de múltiplas camadas.

O neurônio em si calcula uma função muito simples, no qual um certo número de valores de entrada são recebidos, processados e passados para outros neurônios como o valor de saída. No entanto, quando organizados numa estrutura de rede altamente interligada, a função global da rede torna-se muito mais sofisticada e capaz de realizar tarefas computacionais complexas, sendo capazes de realizar mapeamento não-linear de entrada-saída, generalização, adaptatividade e tolerância a falhas (LIN; LEE, 1996).

De um ponto de vista operacional, as redes neurais podem ajustar a sua estrutura durante a fase de aprendizagem de tal modo a adquirirem a capacidade de generalizar objetos diferentes dos dados de treinamento. Por outro lado, a principal desvantagem das redes neurais é a difícil compreensão de como elas realmente resolvem um

determinado problema. Isso porque o problema não é dividido em seus elementos lógicos, mas é resolvido por meio de uma abordagem holística, o que pode ser de difícil compreensão, principalmente, aos leigos nesse tipo de solução.

3.2 Sistemas difusos

A Teoria dos Conjuntos Difusos (do inglês, *Fuzzy Sets Theory*), formalizada pelo professor Lofti Zadeh na Universidade da Califórnia em 1965 (ZADEH, 1965), foi desenvolvida para tratar do aspecto vago da informação. Ela constitui o ponto de partida no desenvolvimento de sistemas difusos. Na publicação de Zadeh (1965), se introduz, pela primeira vez de maneira formal, a definição de um conjunto difuso. Esta dá origem a uma série de conceitos, operações e medidas que são aplicáveis a várias disciplinas da ciência.

Esta seção apresenta as definições básicas, notações e operações que englobam os conjuntos difusos. Isso com o objetivo de fornecer uma introdução concisa e um resumo dos conceitos básicos e fundamentais para o estudo dos conjuntos difusos.

3.2.1 Conjuntos clássicos

Um conjunto clássico é um conjunto com uma fronteira nítida – também conhecido como *crisp*. Por exemplo, um conjunto clássico de um número real maior do que 1,8 pode ser expresso como

$$A = \{x \mid x > 1,8\}, \quad (3.1)$$

onde 1,8 é uma fronteira clara e inequívoca tal que se x é maior do que esse número, então x pertence ao conjunto A ; caso contrário x não pertence ao conjunto.

Embora conjuntos clássicos sejam adequados para diversas aplicações e tenham provado serem uma ferramenta importante para a matemática e ciência da computação, eles não refletem a natureza dos conceitos e pensamentos humanos, que tendem a ser abstratos e imprecisos. Como ilustração, matematicamente podemos expressar o conjunto de pessoas altas como uma coleção de pessoas cuja altura é de mais de 1,80 m. Este é o conjunto denotado pela Equação 3.1, se tomarmos $A =$ “pessoa alta” e $x =$ “altura”. No entanto, esta é uma forma artificial e inadequada de representar o nosso conceito usual de “pessoa alta”, já que a natureza dicotômica do conjunto clássico classificaria uma pessoa com 1,81 metros de altura como alta, mas não uma pessoa de 1,79 metros de altura. Esta distinção é intuitivamente razoável, mas a falha vem da transição brusca entre pertencimento e não pertencimento em um conjunto.

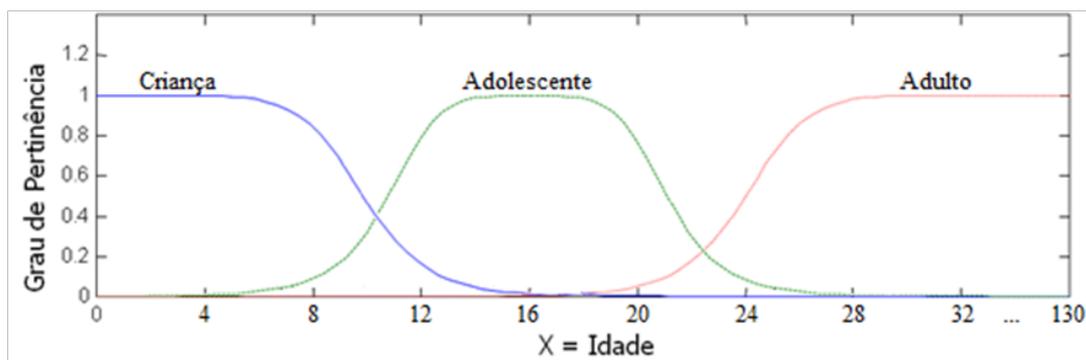


Figura 3.3 - Partição difusa do universo X representando “idade”, formada pelos conjuntos difusos “criança”, “adolescente” e “adulto”.

Outro exemplo de problema que mostra um limitante da modelagem de conjuntos clássicos é a modelagem do conceito de adolescente. Em conjuntos clássicos o conceito de adolescente poderia ser definido como a faixa etária de pessoas entre 13 e 19 anos. Assim, uma pessoa de 12 anos e 11 meses de idade não poderia ser considerada adolescente, o que é incompatível com o raciocínio humano, já que o ser humano admite um grau de compatibilidade intermediário entre o pertencimento ou não de um elemento a um certo conjunto. Uma modelagem difusa admite um grau de pertinência para a afirmação de que uma pessoa de 12 anos e 11 meses de idade é, até certo ponto, um adolescente. A Figura 3.3 mostra uma modelagem para os conjuntos difusos “criança”, “adolescente” e “adulto”.

Em contraste com um conjunto clássico, um conjunto difuso, como o nome indica, é um conjunto sem uma fronteira nítida. Ou seja, a transição de “pertencer a um conjunto” para “não pertence a um conjunto” é gradual, e essa transição suave é caracterizada por funções de pertinência que dão aos conjuntos difusos flexibilidade na modelagem de expressões linguísticas comumente utilizadas, tais como “a água está quente” ou “a temperatura é alta”. Como destacado em Zadeh (1965), tais conjuntos mal definidos desempenham um papel importante no pensamento humano, particularmente nos domínios do reconhecimento de padrões, comunicação de informações e abstração. Note que a imprecisão não vem da aleatoriedade dos membros constituintes dos conjuntos, mas da natureza incerta e imprecisa de pensamentos e conceitos abstratos.

3.2.2 Conceitos básicos da teoria dos conjuntos difusos

Seja X um espaço de objetos e x um elemento genérico de X . Um conjunto clássico A , tal que $A \subseteq X$, é definido como uma coleção de elementos ou objetos $x \in X$, de tal forma que cada x pode pertencer ou não ao conjunto A . Ao definir uma função característica para cada elemento x em X , podemos representar um conjunto clássico por um conjunto de pares ordenados $(x, 0)$ ou $(x, 1)$, o que indica $x \in A$ ou $x \notin A$, respectivamente.

Ao contrário do conjunto convencional anteriormente mencionado, um conjunto difuso expressa o grau em que um elemento pertence ao conjunto. Assim a função característica de um conjunto difuso pode conter valores reais entre 0 a 1, o que denota o grau de pertinência de um elemento qualquer em um determinado conjunto.

Segundo Zadeh (1965), um conjunto difuso é definido como: “Uma classe de objetos com graus de pertinência contínuos. Tal conjunto é caracterizado por uma função de pertinência que varia entre 0 e 1”. De acordo com esta definição, um conjunto difuso A está composto por duas partes essenciais: seus elementos x pertencentes ao universo de discurso X e a função de pertinência μ_A que assina o grau de pertinência destes elementos a seu conjunto. Um conjunto difuso A é definido matematicamente como:

$$\mu_A : X \rightarrow [0, 1]. \quad (3.2)$$

Desta forma, na representação de um conjunto difuso, a função de pertinência associa a cada elemento $x \in X$ um número real $\mu_A(x)$ no intervalo $[0, 1]$, que representa o grau de pertinência do elemento x ao conjunto A , isto é, qual a compatibilidade do elemento x com o conjunto A . A função de pertinência $\mu_A(x)$ indica o grau de compatibilidade entre x e o conceito expresso por A :

- $\mu_A(x) = 0$ indica que x é completamente incompatível com A .
- $0 < \mu_A(x) < 1$ indica que x é parcialmente compatível com A , com grau $\mu_A(x)$.
- $\mu_A(x) = 1$ indica que x é completamente compatível com A .

A especificação da função de pertinência é subjetiva, o que significa que é possível definir mais de uma função de pertinência para o mesmo conceito (digamos, “adoles-

cente”). Isso varia de acordo com a percepção do conceito da pessoa que a define. Por isso, Zimmermann (2001) considera a função de pertinência o componente principal de um conjunto difuso.

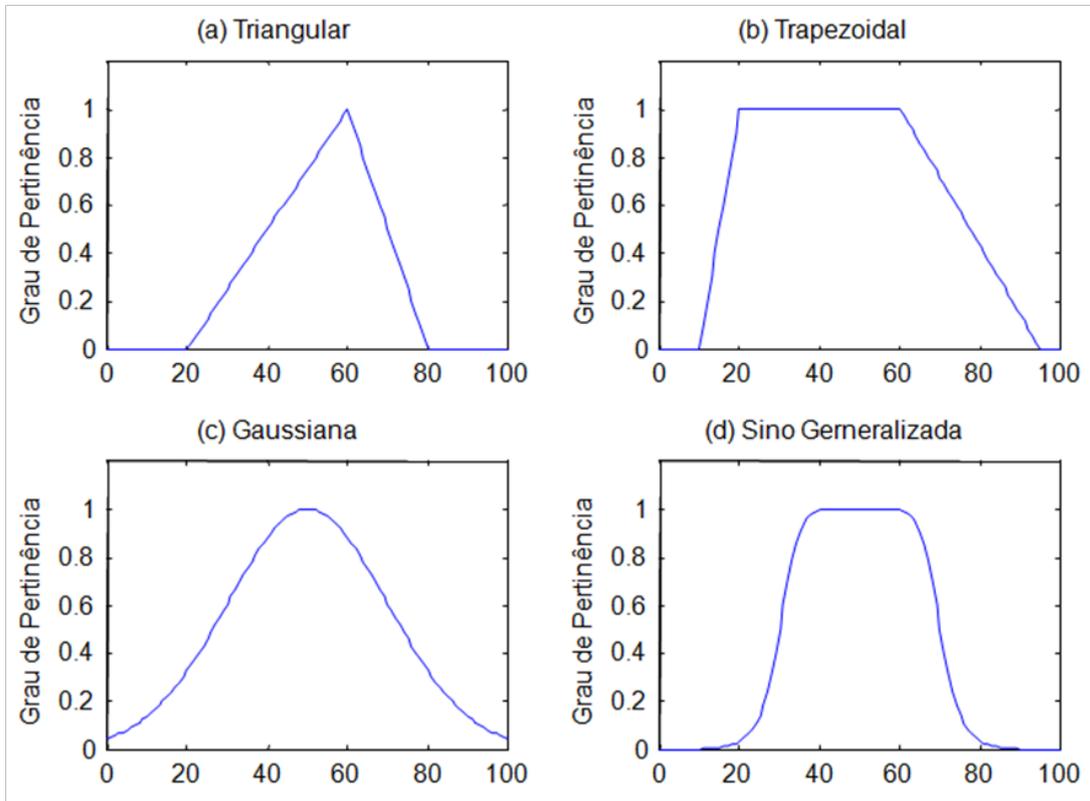


Figura 3.4 - Funções de pertinência.

O tipo de representação da função de pertinência depende do conjunto base. Se este conjunto é composto de muitos valores então uma representação paramétrica é adequada. Assim, neste caso, para que as funções sejam usadas necessita-se apenas que parâmetros sejam definidos. Normalmente funções de pertinência lineares são preferidas, devido à sua simplicidade e eficiência com respeito a computabilidade, principalmente funções trapezoidais ou triangular, que são definidas por quatro e três parâmetros, respectivamente. Entretanto, para algumas aplicações, a modelagem requer curvas continuamente diferenciáveis e transições suaves, que os trapézios não têm (veja Figura 3.4).

3.2.2.1 Operadores de conjuntos difusos

Uma das consequências naturais à generalização da teoria clássica dos conjuntos para a teoria dos conjuntos difusos é a definição de suas operações. Assim como os conjuntos clássicos possuem operações, como união, interseção e complemento, os conjuntos difusos têm operações semelhantes. A seguir, apresentam-se as operações principais com conjuntos difusos, encontradas na literatura consolidada da área (vide, por exemplo, Dubois e Prade (2012) e Gomide et al. (1995)).

Operadores de Intersecção, União e Complemento

Sejam A e B conjuntos definidos em X , tanto em conjuntos difusos como em conjuntos clássicos, é possível expressar a intersecção destes conjuntos, como um outro conjunto E , tal que $E = A \cap B$. Da mesma forma, é possível expressar a união de A e B como um novo conjunto F , tal que $F = A \cup B$. Por outro lado, como visto anteriormente, nos conjuntos difusos o grau de pertinência de um elemento a um conjunto não é binário, por isso os conjuntos resultantes das operações de intersecção e união são dados por funções difusas que as qualificam. Essas funções são geralmente referidas na literatura como t -normas e t -conormas (ou S -normas), respectivamente. Verifica-se que quando os conjuntos são clássicos as t -normas e t -conormas se reduzem aos operadores clássicos e e ou :

$$\textbf{Intersecção:} \textit{ Seja } x \in A \cap B, \textit{ entao } x \in A \textit{ e } x \in B, \quad (3.3)$$

$$\textbf{União:} \textit{ Seja } x \in A \cup B, \textit{ entao } x \in A \textit{ ou } x \in B, \quad (3.4)$$

Em conjuntos difusos, existem diversos operadores t -norma e t -conorma, que mapeiam $\nabla : [0, 1]^2 \rightarrow [0, 1]$, sendo os principais os contidos na Tabela 3.1, e os mais utilizados os operadores de Zadeh e probabilistas. Uma característica comum a todos eles é que eles são:

- Comutativos: $\nabla(a, b) = \nabla(b, a)$;
- Associativos: $\nabla(a, \nabla(b, c)) = \nabla(\nabla(a, b), c)$;
- Monotônicos: $\nabla(a, b) \leq \nabla(c, d)$ se $a \leq c$ e $b \leq d$;

para $\forall a, b \in [0, 1]$. Além disso, os operadores t -norma \top devem verificar a

propriedade de *elemento neutro* = 1, tal que $\top(a, 1) = a$, e os operadores *t-conorma* \perp a propriedade de *elemento neutro* = 0, tal que $\perp(a, 0) = a$.

Uma *t-norma* \top e uma *t-conorma* \perp são duais em relação a uma operação de negação $\neg : [0, 1] \rightarrow [0, 1]$ se elas satisfazem as relações de De Morgan, dadas por:

$$\neg(\top(a, b)) = \perp(\neg a, \neg b); \quad (3.5)$$

$$\neg(\perp(a, b)) = \top(\neg a, \neg b). \quad (3.6)$$

Tabela 3.1 - Principais *t-normas* e *t-conormas* duais.

Nome	<i>t-norma</i>	<i>t-conorma</i>
Zadeh	$\min(a, b)$	$\max(a, b)$
probabilista	$a \cdot b$	$a + b - a \cdot b$
Lukasiewicz	$\max(a + b - 1, 0)$	$\min(a + b, 1)$
Weber	a , se $b = 1$ b , se $a = 1$ 0, senão	a , se $b = 0$ b , se $a = 0$ 0, senão

Operadores de implicação

Os operadores de implicação $I : [0, 1]^2 \rightarrow [0, 1]$ são usados para modelar regras de inferência do tipo *Se <premissa> Então <conclusão>*. Esses operadores devem satisfazer, pelo menos, as três seguintes propriedades $\forall a, a', b, b' \in [0, 1]$:

- a) Se $b \leq b'$ então $I(a, b) \leq I(a, b')$ (monotonicidade à direita);
- b) $I(0, b) = 1$ (princípio da falsidade);
- c) $I(1, b) = b$ (princípio da neutralidade).

Considerando A e B dados por $\mu_A : X \rightarrow [0, 1]$ e $\mu_B : Y \rightarrow [0, 1]$, a relação $A \rightarrow B$ é expressa como:

$$\mu_{A \rightarrow B}(x, y) = I(\mu_A(x), \mu_B(y)), \quad (3.7)$$

onde I é algum operador de implicação. A Tabela 3.2 mostra os principais operadores de implicação disponíveis na literatura.

Tabela 3.2 - Principais operadores de implicação.

Nome	Implicação
Kleene-Dienes	$\max(1 - a, b)$
Lukasiewicz	$\min(1 - a + b, 1)$
Rescher-Gaines “Sharp”	1, se $a \leq b$ 0, senão
Brower-Gödel	1, se $a \leq b$ b , senão
Goguen	$\min(b/a)$, se $a \neq 0$ 1, senão
Reichenbach “Estocástica”	$1 - a + a \cdot b$
Zadeh-Wilmott	$\max(1 - a, \min(a, b))$

Embora as t – normas não sejam implicações propriamente ditas, elas são muito empregadas na prática como tais (SANDRI; CORREA, 1999).

3.2.2.2 Modelos de inferência difusos

Sistemas de inferência difusos são implementações de modelos capazes de processar de forma eficiente informações imprecisas e qualitativas, sendo, por exemplo, adequados em situações que exigem tomadas de decisão (REZENDE, 2003).

Os sistemas de inferência difusos utilizam-se de regras lógicas para descrever, numa rotina, a experiência humana, intuição e heurística para inferir um valor de saída a partir de elementos de entrada, como, por exemplo, num processo de classificação.

A base de regras é formada por estruturas do tipo *Se <premissa> Então <conclusão>*. A premissa pode ser formada por mais de uma cláusula $x_i = A_i$, onde A_i é um conjunto difuso, conectados por operadores de conjunção ou disjunção (t – normas e t – conormas). Essas regras, juntamente com os dados de entrada, são processados pelo procedimento de inferência, o qual infere valores de saída de acordo com o estado do sistema, aplicando o operador de implicação.

Os tipos de modelos de inferência difusos encontrados na literatura são os modelos clássicos, principalmente o modelo de Mamdani e o de Larsen, e os modelos de interpolação, como o modelo de Takagi-Sugeno-Kang (TSK) e o de Tsukamoto (MAMDANI, 1976; LEE, 1990). Os modelos diferem quanto à forma de representação dos termos na premissa, quanto à representação da conclusão e quanto aos operadores utilizados.

Sandri e Correa (1999) definem o processo de inferência como:

- a) Verificação do grau de compatibilidade entre os fatos e as cláusulas nas premissas das regras;
- b) Determinação do grau de compatibilidade global da premissa de cada regra, utilizando os operadores de conjunção e disjunção (*t – normas* e *t – conormas*);
- c) Determinação do valor da conclusão mediante a aplicação de função de implicação, tendo como argumentos o grau de compatibilidade da regra com os dados e o conseqüente da regra (preciso ou não);
- d) Agregação dos valores obtidos como conclusão nas várias regras, obtendo-se um resultado global.

Modelo clássico

Nos modelos clássicos, a conclusão de cada regra especifica um termo difuso dentre um conjunto fixo de termos. Esses termos são geralmente conjuntos difusos convexos como triângulos, funções em forma de sino e trapézios.

Dado um conjunto de valores para as variáveis de estado, o sistema obtém um conjunto difuso – muitas vezes sub-normalizado –, como o valor da variável de saída. Esse conjunto difuso representa uma ordenação no conjunto de saídas aceitáveis naquele momento. Finalmente, uma saída global é selecionada dentre aquelas aceitáveis em um processo conhecido como decodificação (SANDRI; CORREA, 1999). Na Figura 3.5, ilustra-se o processo de raciocínio do modelo de Mamdani, para duas regras com duas variáveis de entrada, que usa o operador *min* como implicação e o operador *max* para agregar os resultados de cada regra.

Modelo interpolativo

Nos modelos de interpolação, cada conclusão é dada através de uma função estritamente monotônica, usualmente diferente para cada regra. No modelo de Takagi-Sugeno-Kang, a função é uma combinação linear das entradas, tendo como parâmetros um conjunto de constantes. Nesse esquema, obtém-se, para cada regra, um único valor para a variável de saída. Finalmente, uma saída global é obtida fazendo-se uma média ponderada dos valores individuais obtidos, onde cada peso é o próprio grau de compatibilidade entre a premissa da regra e as entradas, normalizado. Na Figura

3.5, ilustra-se a inferência através do método de Takagi-Sugeno-Kang para as mesmas duas regras e duas variáveis de entrada que o exemplo de Mamdani mencionado anteriormente.

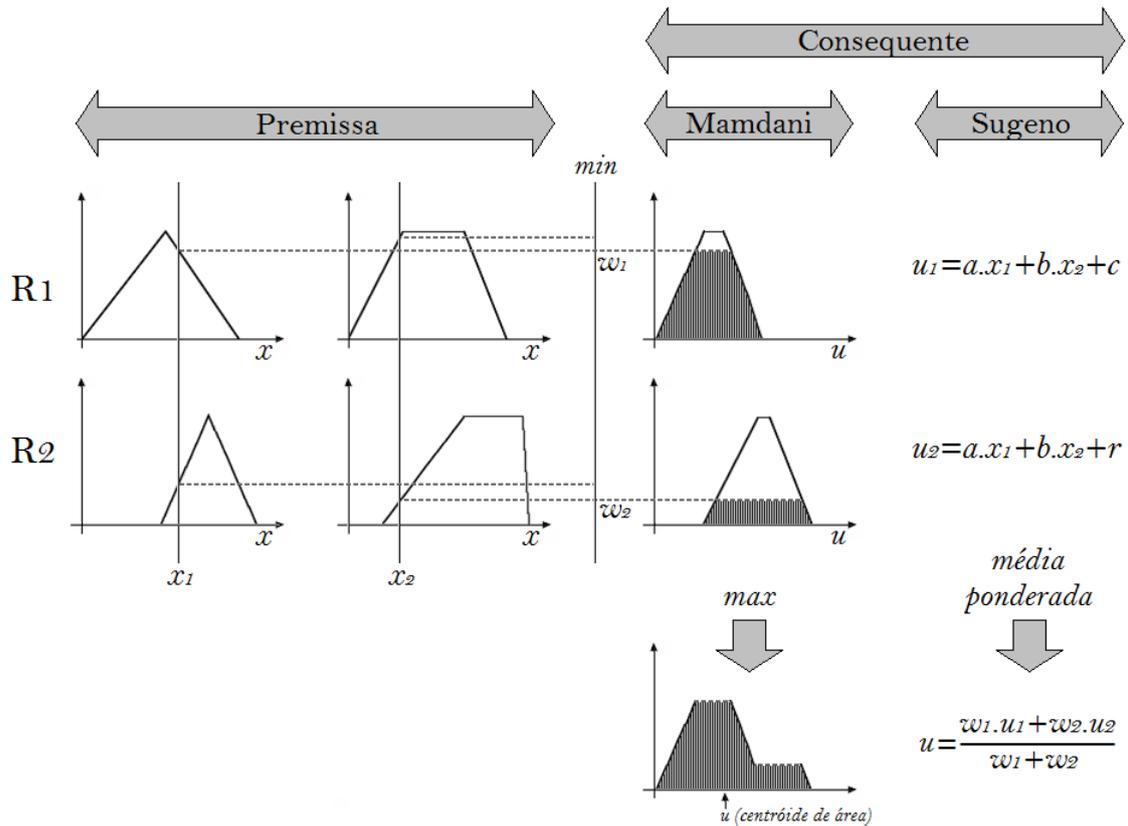


Figura 3.5 - Duas regras SE-ENTÃO (R1 e R2), com duas funções de pertinência de variáveis de interesse, e métodos de inferência difusos comumente utilizados na literatura.

3.3 Sistemas neuro-difusos

A lógica difusa (ZADEH, 1965; RUSPINI et al., 1998) e as redes neurais artificiais (HAYKIN, 2001; MEHROTRA et al., 1996) são tecnologias complementares no projeto de sistemas inteligentes. A combinação destas duas tecnologias em um sistema integrado, leva a criação de sistemas capazes de captar algumas das qualidades que caracterizam o cérebro humano.

As redes neurais são, essencialmente, algoritmos computacionais de baixo nível que, usualmente, oferecem um bom desempenho nas tarefas de reconhecimento de pa-

drões e controle. Elas podem aprender a partir de conjuntos de dados, enquanto as soluções de lógica difusa são fáceis de verificar e de otimizar. O sistema integrado tende a ter as vantagens das duas técnicas:

- a) das redes neurais, oferece as habilidades de aprendizagem, de otimização e a estrutura conexionista;
- b) dos sistemas difusos, oferece o estilo de raciocínio humano e a facilidade de incorporação de conhecimento especializado.

Desta forma, é possível trazer a aprendizagem de baixo nível e o poder computacional de redes neurais em sistemas difusos e também o estilo de pensamento humano de alto nível SE-ENTÃO de sistemas difusos dentro de redes neurais.

Existem diversas maneiras de combinar redes neurais e lógica difusa. Os esforços de fusão dessas duas técnicas podem ser caracterizados considerando três categorias principais: sistemas difusos neurais, sistemas neurais difusos e sistemas híbridos neuro-difusos (CANUTO, 2001). Estes três sistemas estão detalhados a seguir:

- **Sistemas difusos neurais**

Sistemas difusos neurais são caracterizados pela utilização de redes neurais para fornecer sistemas difusos com um tipo de método de ajuste automática de parâmetros. Um exemplo dessa abordagem seria a utilização de redes neurais para a obtenção das funções de pertinência utilizados nas regras difusas. Este tipo de combinação é usado principalmente em aplicações de controle.

Um exemplo desta abordagem pode ser encontrado no trabalho proposto por Jang (1993), o ANFIS (do inglês, *Adaptative NeuroFuzzy Inference System*).

O sistema de inferência adaptativo neuro-difuso (ANFIS, do inglês *Adaptive Neuro Fuzzy Inference System*), utilizando um conjunto de dados entrada/saída, constrói um sistema de inferência difuso, do modelo TSK, cujos parâmetros da função de pertinência são ajustados utilizando um algoritmo de retropropagação padrão ou combinado a algum método como o dos mínimos quadrados. Isso permite que seus sistemas difusos aprendam com os dados que estão sendo modelados (JANG, 1993). A Figura 3.6 mostra o esquema do modelo ANFIS.

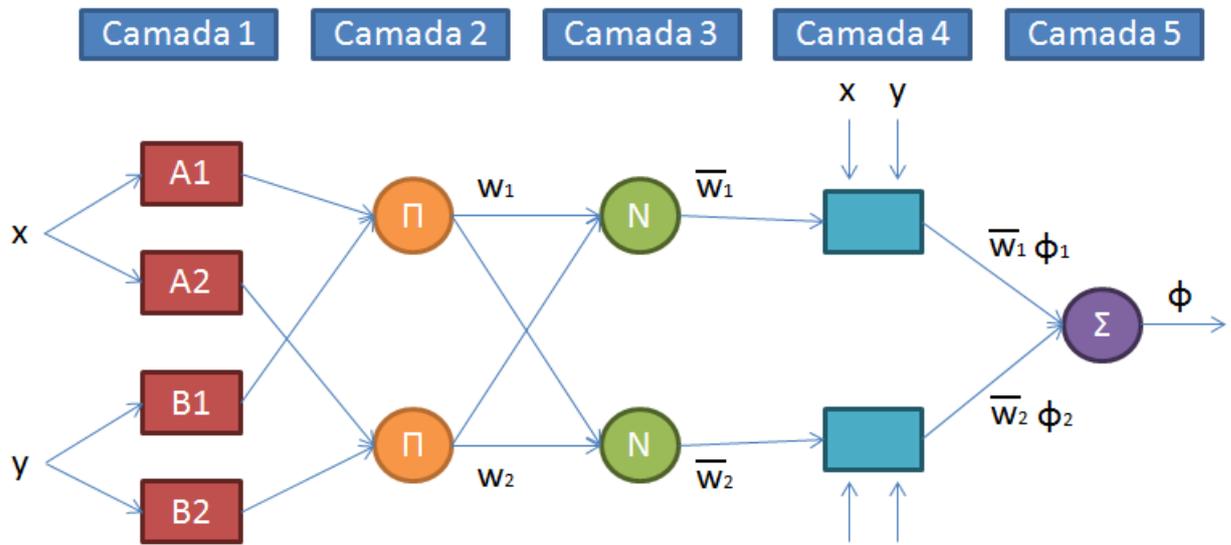


Figura 3.6 - Sistema ANFIS.

O modelo apresentado na Figura 3.6 é formado por cinco camadas, as quais serão descritas a seguir:

- Camada 1 - Sendo x e y os valores de entrada para o sistema, os nós desta camada ($A1$, $A2$, $B1$ e $B2$) representam o grau de pertinência destes nós em relação aos valores de entrada, ou seja, esta camada realiza a fase de codificação do sistema.
- Camada 2 - As saídas dos nós desta camada representam a força de disparo de uma regra. Qualquer operador de t -norma que represente o operador E lógico pode ser utilizado como função de ativação nesta camada (JANG, 1993).
- Camada 3 - Cada nó desta camada funciona como um operador de normalização do nível de disparo das regras.
- Camada 4 - Esta camada calcula o produto entre as saídas da camada 3 (níveis de disparo normalizados) e os valores de entrada x e y , ponderados, os quais Jang (1993) chamou de parâmetros consequentes.
- Camada 5 - Esta camada realiza a fase de decodificação do sistema. É formada por apenas um nó que realiza o somatório das saídas da camada 4.

O método neuro-adaptativo de aprendizagem funciona de forma semelhante a das redes neurais. Técnicas neuro-adaptativas de aprendizagem

proporcionam um método para o procedimento de modelagem difusa para aprender informações sobre um conjunto de dados. O sistema calcula os parâmetros das funções de pertinência que melhor permitem o sistema de inferência difuso rastrear os dados de entrada/saída.

- **Sistemas neurais difusos**

O objetivo principal desta abordagem é o de inserir modificações difusas em algum, ou alguns, dos elementos de redes neurais, utilizando a lógica difusa e a teoria dos conjuntos difusos. Neste caso, um neurônio convencional pode tornar-se difuso. Os sistemas neurais difusos são inerentemente redes neurais, que são utilizados principalmente em aplicações de reconhecimento de padrões.

Em [Lin e Lee \(1996\)](#), por exemplo, uma rede neural composta de neurônios difusos é apresentado. Nesses neurônios difusos, as entradas são não-difusas, mas as operações de ponderação são substituídas por funções de pertinência. O resultado de cada operação de ponderação é o valor de adesão à entrada correspondente no conjunto difuso.

Neste trabalho, um sistema neural difuso será implementado inserindo-se modificações difusas na rede neural artificial Perceptron Multicamadas. Mais detalhes sobre esse modelo podem ser verificadas no Capítulo 4.

- **Sistemas híbridos neuro-difusos**

Nesta abordagem, as duas técnicas são utilizadas independentemente, tornando-se, nesse sentido, um sistema híbrido. Cada um faz o seu próprio trabalho em diferentes partes do sistema, complementando-se mutuamente, a fim de atingir um objetivo comum. Este tipo de fusão é orientada à aplicação e adequada tanto para aplicações de controle quanto de reconhecimento de padrões ([LIN; LEE, 1996](#)).

4 PERCEPTRON MULTICAMADAS DIFUSO

Este capítulo apresenta uma versão modificada difusa do modelo perceptron de múltiplas camadas que usa, como base, a implementação desenvolvida em Pal e Mitra (1992). A adição das modificações difusas no MLP (do inglês *Multilayer Perceptron*) tradicional foi necessária devido ao alto grau de complexidade dos padrões de treinamento, principalmente pela resolução espacial – que permite a existência de mais de uma classe de uso da terra num mesmo pixel – e a ausência da verdade terrestre. Assim, uma solução com a incorporação de lógica difusa se tornou mais atraente por combinar o estilo de raciocínio humano de sistemas difusos e a estrutura conexionista e de aprendizagem de redes neurais. O modelo desenvolvido faz uso de uma função de pertinência que associa valores de similaridade de um padrão de entrada às classes do problema, gerando um vetor difuso de saída desejada. Além disso, adotamos um parâmetro na equação de atualização dos pesos das conexões a fim de levar em conta o grau de ambiguidade de um padrão de entrada durante a fase de aprendizagem.

4.1 Introdução

O perceptron multicamadas, ou perceptron de múltiplas camadas, utilizando o mecanismo de aprendizagem de retropropagação, é a rede neural mais comum encontrada na literatura (MEHROTRA et al., 1996), e tem sido utilizado numa vasta gama de aplicações, tais como reconhecimento de padrões (DIMLA; LISTER, 2000; JEONG et al., 2000; ZHANG, 1999), análise de imagem médica (GÜLER et al., 1998; SHEPPARD et al., 1999) e de previsão (INDRO et al., 1999). Essencialmente, o MLP é uma rede *feed-forward* multicamadas que utiliza um mecanismo de aprendizagem supervisionada baseada no ajuste de seus parâmetros de acordo com o erro entre as saídas desejada e real da rede. Ou seja, ele utiliza um mecanismo que modifica os pesos da rede, a fim de minimizar o erro quadrático médio entre as saídas desejadas e reais da rede (RUMELHART et al., 2002).

O perceptron multicamadas difuso é uma implementação de uma modificação, baseada na teoria dos conjuntos difusos, em um MLP clássico (PAL; MITRA, 1992; KELLER; HUNT, 1985). Em outras palavras, é o resultado da fuzzificação direta quer a um nível de rede, de aprendizagem ou de rede e aprendizagem do perceptron multicamadas.

Nesta seção, fundamentamos a implementação de uma fuzzificação em nível de rede e outra em nível de aprendizagem, além da combinação dessas duas técnicas para

formar uma fuzzificação em nível de rede e aprendizagem. O modelo MLP difuso desenvolvido utiliza como base o trabalho de [Pal e Mitra \(1992\)](#). As modificações da rede MLP são as seguintes:

- **Saída desejada difusa:** Neste mecanismo, são empregados conceitos difusos, a fim de calcular a saída desejada de padrões, os quais são apresentados para a rede MLP durante a fase de aprendizagem;
- **Grau de ambiguidade:** Neste mecanismo, um parâmetro que leva em conta o grau de ambiguidade do padrão de entrada foi adicionado à equação de atualização dos pesos das conexões.

Este capítulo está dividido da seguinte forma. Primeiramente, uma breve descrição do perceptron multicamadas convencional é apresentado. Em seguida, é apresentada uma descrição do perceptron multicamadas difuso proposto em [Pal e Mitra \(1992\)](#).

4.2 A rede perceptron multicamadas

O modelo usado neste trabalho foi o MLP, que possui aprendizado supervisionado, por correção de erros e atualização dos pesos sinápticos das conexões baseado no algoritmo de retropropagação ([HAYKIN, 2001](#)).

Na fase de treinamento um conjunto de padrões de entrada é apresentado à rede e, por meio de cálculos que incluem sinais sinápticos e pesos de conexões, ele é propagado da camada de entrada até a camada de saída gerando, nesta última, um vetor de saída calculada da rede.

A diferença entre a saída calculada pela rede e a saída real do padrão apresentado é o erro que é propagado para trás, contra a direção das conexões sinápticas. Os pesos são ajustados para que a resposta obtida pela rede convirja, em um sentido estatístico, para a resposta real apresentada no padrão.

Nesta implementação, a camada de entrada é um conjunto de neurônios de mesma quantidade que o tamanho do padrão de entrada. Ou seja, se existem N valores neste conjunto – desprezando-se a resposta desejada – serão criados N neurônios na camada de entrada, sendo que cada um receberá um dos valores do padrão, para cada padrão apresentado. Similarmente, a camada de saída possui a mesma quantidade de neurônios que o número de elementos do vetor de resposta desejada.

Existem apenas conexões entre neurônios de camadas adjacentes, sendo que cada

neurônio de uma camada possui conexões com todos os neurônios das camadas subsequentes. Portanto, entre duas camadas vizinhas tem-se a mesma propriedade de um grafo bipartido completo (ver Figura 4.2).

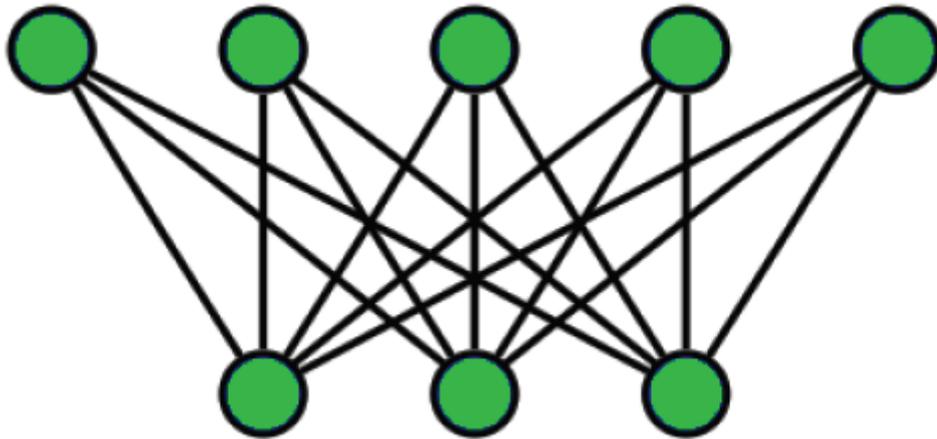


Figura 4.1 - Exemplo de grafo bipartido completo.

4.2.0.3 Função de ativação

No perceptron multicamadas cada neurônio possui uma função de ativação contínua e não linear, que tem por objetivo simular a frequência dos potenciais de ação, ou disparo, de neurônios biológicos no cérebro. Esta função pode ser modelada de várias formas, mas deve ser sempre normalizável e diferenciável.

De acordo com [Haykin \(2001\)](#) existem três tipos básicos de funções de ativação:

- Função de Limiar: É uma função degrau que assume os valores 0, desativado, ou 1 ativado.
- Função Linear por Partes: É uma evolução da função de limiar e possui a propriedade de ser linear até que se atinja os limiares de saturação máximo e mínimo, quando a saída torna-se constante. Essa é uma função contínua, entretanto não diferenciável.
- Função Sigmoides: Possui um gráfico em forma de ‘S’ e é a função mais utilizada como função de ativação na construção de redes neurais artificiais. Ela é uma função estritamente crescente que mostra um comportamento

adequado e intermediário entre linear e não linear. Um exemplo de função sigmoide é a função logística, definida por

$$y_i = \varphi(v_i) = (1 + e^{-a \cdot v_i})^{-1}, \quad (4.1)$$

onde a é o parâmetro de inclinação da função sigmoide, neste trabalho assume o valor 1, y_i é a saída do neurônio i e v_i é a soma ponderada das sinapses de entrada deste neurônio.

4.2.0.4 O processo de treinamento da rede

No MLP o “conhecimento” da rede fica armazenado nos pesos das conexões e, assim, a partir do momento que a rede estiver treinada, podem-se avaliar novos casos sem que para isso se precise do parecer de um especialista.

Basicamente, a aprendizagem por retropropagação de erro consiste em dois passos (HAYKIN, 2001): um passo para frente, a propagação, e um passo para trás, a retropropagação.

Na propagação, um padrão de atividade x (vetor de entrada) de tamanho N é aplicado aos neurônios sensoriais da rede (camada de entrada). Nessa camada não há cálculo da função de ativação, assim tomamos a saída de seus neurônios como $y_i = x_i$, sendo $i = (1, 2, \dots, N)$ o índice de cada neurônio da camada.

Os neurônios da camada seguinte (primeira camada escondida) calculam a soma v_i do produto das saídas y_j da camada anterior pelos pesos w_{ij} das conexões que unem a camada atual e a camada anterior. Nesse momento $i = (1, 2, \dots, N_k)$ é o índice de cada neurônio da camada atual k , de tamanho N_k , e $j = (1, 2, \dots, N_{k-1})$ é o índice dos neurônios da camada anterior $k - 1$. Assim

$$v_i = \sum_{j=1}^n y_j \cdot w_{ij}. \quad (4.2)$$

A força de disparo do neurônio i é

$$y_i = \varphi(v_i) = (1 + e^{-v_i})^{-1}. \quad (4.3)$$

Esse processo é repetido camada por camada até a última camada, onde um conjunto

de saída é produzido como a resposta calculada da rede.

Por outro lado, na retropropagação, os pesos sinápticos são todos ajustados de acordo com uma regra de correção de erro. Especificamente, a resposta calculada da rede é subtraída de uma resposta desejada para produzir o sinal de erro:

$$e_i(n) = d_i(n) - y_i(n), \quad (4.4)$$

sendo $d_i(n)$ a saída desejada na iteração n , isto é, na apresentação do n -ésimo vetor de treinamento.

Desse modo, o valor instantâneo da soma dos erros quadráticos, na camada de saída, e a energia média desse erro, respectivamente, são:

$$\xi(n) = \frac{1}{2} \sum_{i=1}^{N_k} e_i(n)^2 \quad (4.5)$$

e

$$\xi_{med}(n) = \frac{1}{N} \sum_{n=1}^N \xi(n). \quad (4.6)$$

O cálculo da atualização de cada peso, usando gradiente descendente, é dado pela seguinte equação:

$$\Delta w_{ij}(n) = -\eta \cdot \frac{\partial \xi(n)}{\partial w_{ij}(n)} \quad (4.7)$$

ou

$$\Delta w_{ij}(n) = \eta \cdot \delta_i(n) \cdot y_i, \quad (4.8)$$

onde o gradiente local é $\delta_i(n) = \frac{\partial \xi(n)}{\partial v_i(n)} = e_i(n) \cdot \varphi'(v_i(n))$, sendo φ' a derivada da função de ativação descrita anteriormente e η a taxa de aprendizagem, que deve ser cuidadosamente selecionada para garantir que os pesos convirjam rápida e suavemente. Na programação de aplicações, esse parâmetro geralmente é situado entre

0.2 e 0.8.

A análise é mais difícil de ser feita para as mudanças que ocorrem nas camadas ocultas, mas é possível demonstrar que a derivação relevante é:

$$\delta_i(n) = \varphi'(v_i(n)) \cdot \sum_{k=1}^K \delta_k(n) \cdot w_{ik}(n). \quad (4.9)$$

Isso depende da atualização dos pesos da k -ésima camada, que representa a camada posterior à camada atual. Então, a atualização dos pesos das conexões começa entre a camada de saída e a camada anterior e continua camada por camada até a entrada da rede.

Com o objetivo de acelerar a convergência e garantir a estabilidade é usual adicionar ao algoritmo de treinamento o coeficiente *momentum* α , para $0 \leq \alpha \leq 1$.

Este método consiste em adicionar ao cálculo do valor da mudança do peso sináptico uma fração proporcional à alteração anterior. Assim, a introdução desse termo na equação de adaptação dos pesos tende a acelerar a convergência da rede e aumentar a estabilidade do processo de aprendizado, favorecendo mudanças na mesma direção. A Equação 4.10 especifica o ajuste das conexões entre unidades de processamento pela aplicação do termo *momentum*.

$$\Delta w_{ij}(n) = \eta \cdot \delta_i(n) \cdot y_j + \alpha \cdot \Delta w_{ij}(n-1). \quad (4.10)$$

4.3 O perceptron multicamadas difuso

Esta seção apresenta uma versão modificada difusa do modelo perceptron de múltiplas camadas que usa, como base, a implementação desenvolvida em Pal e Mitra (1992). Apesar de o perceptron multicamadas tradicional ser uma solução aparentemente suficiente para resolver o nosso problema de forma satisfatória, uma solução híbrida – com a incorporação de lógica difusa – se tornou mais atraente por combinar o estilo de raciocínio humano de sistemas difusos e a estrutura conexionista e de aprendizagem de redes neurais. O modelo desenvolvido faz uso de uma função de pertinência que associa valores de similaridade de um padrão de entrada às classes do problema, gerando um valor difuso de saída desejada. Além disso, adotamos um parâmetro na equação de atualização dos pesos das conexões a fim de levar em conta o grau de ambiguidade de um padrão de entrada durante a fase de aprendizagem.

4.3.1 Fuzzificação da rede MLP

Como mencionado anteriormente, existe uma vasta literatura sobre a combinação da teoria dos conjuntos difusos (ZADEH, 1965) e o perceptron multicamadas (HAYKIN, 2001; MEHROTRA et al., 1996), tal como em (STOEVA; NIKOV, 2000; PAL; MITRA, 1992; KELLER; HUNT, 1985).

Essencialmente, a combinação destas duas tecnologias pode ser conseguida em três níveis diferentes, que são a fuzzificação em nível de rede, fuzzificação em nível de aprendizagem e fuzzificação em nível de rede e de aprendizagem, como se segue:

- Fuzzificação em nível de rede: Neste método, o processo de fuzzificação ocorre na estrutura do modelo perceptron multicamadas, como em neurônios, pesos e/ou em toda a arquitetura da rede, assim como na saída desejada;
 - Neurônios e arquitetura difusos: Na maioria dos casos, a simulação do comportamento de um sistema difuso no MLP é realizada adicionando conceitos de conjuntos difusos nos neurônios da rede. Outra forma seria o de as camadas de neurônios representarem o conjunto de regras difusas. Nestes casos, um algoritmo de retropropagação convencional é empregado de modo a atualizar e ajustar os parâmetros do sistema difuso. O principal objetivo desta abordagem é a otimização dos parâmetros de sistemas difusos;
 - Saída desejada difusa: Nesta abordagem, o vetor de saída desejada é fuzzificado, a fim de representar o grau de pertinência (similaridade) do padrão de entrada para cada classe de saída (PAL; MITRA, 1992).
- Fuzzificação em nível de aprendizagem: Neste método, o processo de fuzzificação ocorre no algoritmo de aprendizagem, preservando a arquitetura do perceptron multicamadas. O processo de aprendizagem modificado pode ser chamado de retropropagação difuso. Neste caso, o método tenta melhorar alguns pontos fracos da retropropagação tradicional, conseguindo aumentar a taxa de convergência (STOEVA; NIKOV, 2000), melhorando a precisão (HWANG et al., 1997) e diminuindo a influência de padrões de formação ambíguos (KELLER; HUNT, 1985);
- Fuzzificação em nível de rede e aprendizagem: Neste método, o processo de fuzzificação ocorre na estrutura do modelo perceptron multicamadas,

bem como no método de aprendizagem de retropropagação. Ou seja, é a combinação do uso dos dois métodos anteriores.

4.3.1.1 Vetor de saída desejada

Normalmente, no MLP convencional (HAYKIN, 2001; RUMELHART et al., 2002), o número de nós da camada de saída corresponde ao número de classes do conjunto de padrões de entrada. Um método comum de ser utilizado para a saída da rede é o vencedor-leva-tudo, que atribui 1 (um) para o neurônio vencedor e 0 (zero) para os outros neurônios. Dessa forma, o neurônio vencedor representa a predição da rede sobre a classe a que o padrão de entrada pertence.

Em problemas do mundo real, no entanto, os dados são geralmente mal definidos, com sobreposição de classes. Ou seja, podem existir padrões com graus de pertinência diferente de zero a mais de uma classe. A Figura 4.2 é um exemplo simples de padrões que pertencem à região de sobreposição e têm similaridade diferente de 0 (zero) para ambas as classes (A e B).

No perceptron multicamadas convencional, essa semelhança múltipla não é considerada já que apenas o neurônio vencedor define a predição da rede a determinada a classe. A fim de considerar os valores de pertinência de todas as classes, parece muito promissor incorporar os conceitos difusos no cálculo da saída desejada (PAL; MITRA, 1992; CANUTO, 2001).

O perceptron multicamadas difuso pode determinar valores de pertinência desejado – calculado incorporando conceitos difusos – para os neurônios de saída, durante a fase de treinamento, em vez de escolher valores binários como no método vencedor-leva-tudo (PAL; MITRA, 1992). Consequentemente, os erros podem ser propagados de volta em relação à semelhança exata que é refletida na saída desejada (saída desejada difusa).

O uso da saída desejada difusa permite que o modelo perceptron de múltiplas camadas classifique mais eficientemente dados difusos com sobreposição de limites de classes. Além disso, o MLP difuso pode ser aplicado em qualquer tarefa para a qual o MLP convencional é utilizado.

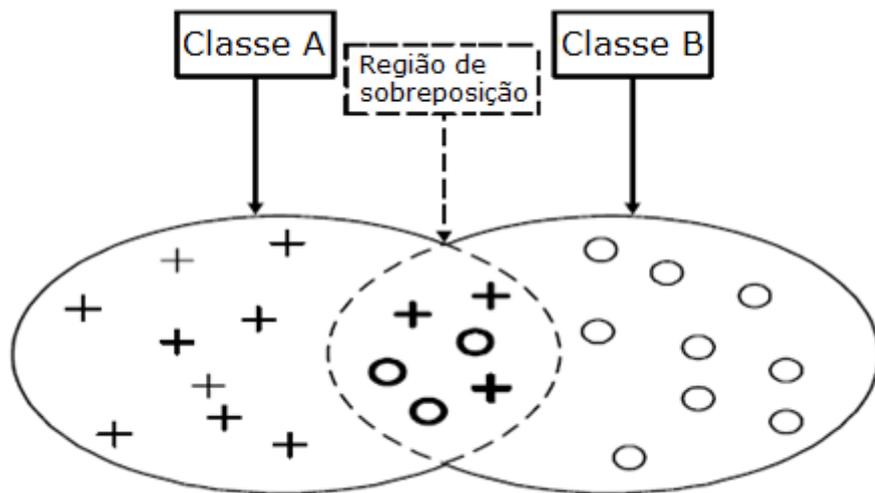


Figura 4.2 - Exemplo de um espaço com duas classes sobrepostas.

Fonte: Adaptado de Canuto (2001).

4.3.1.1.1 Cálculo da saída desejada difusa

Como mencionado antes, ao contrário do perceptron multicamadas convencional, cada saída desejada do perceptron multicamadas difuso se situa no intervalo $[0, 1]$, e refere-se ao grau de pertinência do padrão de entrada para a sua correspondente classe de saída. Em Pal e Mitra (1992), um método para utilizar os conceitos difusos, a fim de calcular a saída desejada, foi proposto. Este método usa como base dois parâmetros:

- O vetor média do conjunto de treinamento;
- Desvio padrão dos padrões de treinamento.

De acordo com Pal e Mitra (1992), a fim de obter a saída desejada difusa, os seguintes passos devem ser executados:

- Calcula-se a distância ponderada entre o padrão de entrada e cada classe ($k = 1, ..C$), levando em conta os vetores média e desvio padrão dos padrões de treinamento. A distância ponderada do padrão de treinamento i para a k -ésima classe é definida por:

$$D_k = \sqrt{\sum_{j=1}^n \left(\frac{x_{ij} - \mu_{kj}}{\sigma_{kj}} \right)^2}, \quad (4.11)$$

em que: x_{ij} é o valor de índice j do padrão de entrada i ; μ_{kj} é o valor de índice j do vetor média da k -ésima classe; σ_{kj} é o valor de índice j do vetor de desvio padrão da k -ésima classe; n é o tamanho do vector padrão.

- Uma vez que a distância ponderada está definida, os valores de pertinência para cada classe (μ_k para $k = 1, ..C$) são calculadas utilizando a seguinte equação:

$$\mu_k(D_k) = \frac{1}{1 + \left(\frac{D_k}{f_d}\right)^{f_e}}, \quad (4.12)$$

em que: D_k é a distância ponderada do padrão de entrada para a classe k ; f_e e f_d são parâmetros difusos.

Estes dois parâmetros difusos (f_e e f_d) controlam a quantidade de imprecisão para esta classe. A ideia por detrás da função de pertinência é que quanto maior a distância de um padrão de entrada em relação a classe k , menos semelhante é o padrão à classe k e, assim, menor será o valor de saída da função de pertinência.

4.3.1.2 O processo de atualização difusa dos pesos

Uma das maneiras de aprendizagem da rede MLP convencional é por meio da minimização do erro médio quadrático (LMS, do inglês *Least Mean Square*) entre o vetor saída desejada e a saída calculada pela rede. Este é conhecido como o método do gradiente descendente. A minimização LMS é conseguida através da atualização dos pesos da rede, como visto na Equação (4.10).

Esta equação atualiza os pesos da rede baseada em seus pesos anteriores de acordo com uma determinada taxa de aprendizagem. O parâmetro α é utilizado nessa equação para ajudar a solução a não ficar presa num mínimo local, caso muito comum em problemas de otimização.

No processo de atualização de pesos apresentado na Equação (4.10), cada padrão de treinamento tem a mesma importância em ajustá-los, no entanto, para padrões que estão em áreas de sobreposição de classes e que são os principais responsáveis por má classificação, essa importância deveria ser diminuída. Na equação de atualização

de pesos, este fator não é considerado e padrões de treinamento ambíguos têm a mesma influência que inequívocos.

Um modo de melhorar o desempenho do algoritmo consiste em considerar a quantidade de correção no vetor peso produzido pelo padrão de entrada. Assim, a quantidade de correção é definida pelo grau de ambiguidade de um padrão, em que quanto mais ambíguo é um padrão, menor é a correção do peso. Portanto, os padrões ambíguos devem ter menos influência no processo de atualização de pesos.

O grau de ambiguidade é um parâmetro adicional para ser usado no perceptron multicamadas difuso, que não foi utilizado no modelo proposto em Pal e Mitra (1992). Esta ideia é semelhante à de Keller e Hunt (1985), que se baseou numa partição de duas classes, mas, proposta por Canuto (2001), é estendida a uma partição de N classes.

O grau de ambiguidade A pode ser definido como se segue.

$$A = (\mu_i(x) - \mu_j(x))^m, \quad (4.13)$$

em que: $\mu_i(x)$ é o valor da classe com maior grau de pertinência para o padrão x ; $\mu_j(x)$ é o valor da classe com segundo maior grau de pertinência para o padrão x ; m é um parâmetro difuso de aumento/redução de ambiguidade.

O parâmetro m aumenta ($m < 1$), mantém ($m = 1$) ou reduz ($m > 1$) a influência da ambiguidade de um padrão de formação, bem como determina a força deste aumento/redução.

O grau de ambiguidade é, então, usado como um parâmetro na equação de atualização de pesos (ver Equação 4.14) no segundo termo da equação, juntamente com a taxa de aprendizagem (η) e a saída do neurônio. A nova equação de atualização de peso passa a ser:

$$\Delta w_{ij}(n) = A \cdot \eta \cdot \delta_i(n) \cdot y_j + \alpha \cdot \Delta w_{ij}(n-1). \quad (4.14)$$

Quando um padrão de entrada é ambíguo, tem alta similaridade (valor de pertinência) com mais de uma classe. Assim, o seu grau de ambiguidade é baixo, diminuindo o nível de correção nos pesos. Em outras palavras, quanto mais um padrão de trei-

namento é ambíguo, menor é a sua influência sobre a equação de atualização pesos.

5 IDENTIFICAÇÃO DE PADRÕES TEMPORAIS DE USO E COBERTURA DA TERRA

Os dados de satélite com alta resolução temporal são muito úteis para a análise do comportamento da variação de atributos da cobertura da terra presente em determinada área de interesse. Esses dados podem revelar padrões sazonais e variações inter-anuais relacionados tanto ao tipo de cobertura do solo quanto à viabilidade natural do clima no ecossistema em questão. Desse modo, o comportamento da variação dos atributos ligados à superfície terrestre pode caracterizar classes de uso do solo e cobertura da terra.

Tradicionalmente, os trabalhos que tratam da comparação entre sucessivos mapas de cobertura da terra estão interessados no monitoramento pontual da conversão da cobertura. Por outro lado, a existência de dados com representação temporal contínua permite a detecção de mudanças sutis dentro de uma classe de cobertura, o que viabiliza uma análise temporal dos dados e não apenas pontual como o habitual. Nesse sentido, mudanças nas características sazonais podem ser relacionadas com mudanças de uso do solo e cobertura da terra (GALFORD et al., 2008), sendo que as variações fenológicas tendem a seguir padrões temporais relativamente bem definidos para cada tipo de cobertura. Na Figura 5.1 são mostrados alguns exemplos de séries temporais do índice de vegetação EVI2 com diferentes características de sazonalidade apresentados por Freitas (2012). Essas curvas mostram o potencial da análise de sazonalidades para o monitoramento da cobertura vegetal.

É possível observar na Figura 5.1 que as sazonalidades podem apresentar grandes variações para diferentes tipos de uso do solo. Um exemplo é a classe floresta, que apresenta flutuações sazonais com pouca diferença de uma curva para outra, mas que, em 2004, (a) visivelmente difere de (b) por conta de uma mudança de cobertura do solo. Sendo que (a), a partir de 2005, mostra uma sazonalidade característica de pastagem. Analisando essa figura fica claro que as diversas classes de uso do solo e cobertura da terra apresentam assinaturas singulares.

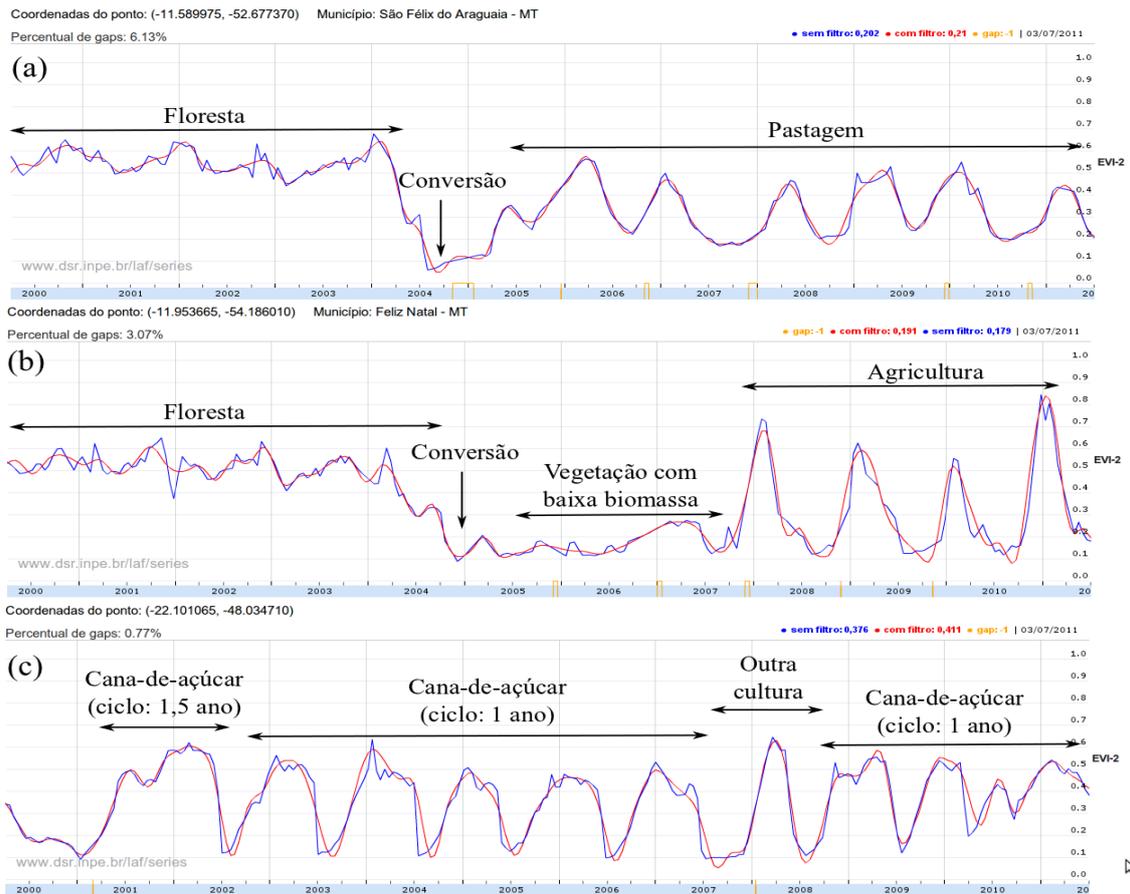


Figura 5.1 - Séries temporais de EVI2 com filtro de wavelet (vermelho) e sem filtro (azul) disponíveis no Laboratório Virtual de Sensoriamento Remoto <<https://www.dsr.inpe.br/laf/series/>>; (a) conversão de floresta para pastagem em 2004; (b) conversão da floresta para outro tipo de vegetação com baixa biomassa em 2004 e em 2007-2008 introdução de agricultura; (c) mudança no sistema de cultivo da cana-de-açúcar com introdução de outra cultura em 2007-2008 e depois novamente cana-de-açúcar.

Fonte: Adaptado de Freitas (2012).

Um estudo sobre intensificação de uso da terra utilizando dados de EVI MODIS foi desenvolvido por Sakamoto et al. (2009), que analisou a expansão da aquicultura e do triplo cultivo de áreas de arroz em uma zona costeira do Vietnã entre os anos de 2000 e 2006. Baseando-se no mesmo princípio do trabalho desenvolvido por Galford et al. (2008) – que utilizou, para identificar áreas de intensificação do uso agrícola, o número de pontos de máximo local –, caracterizaram os tipos de sistema de cultivo de arroz adotados: simples, duplo ou triplo (Figura 5.2).

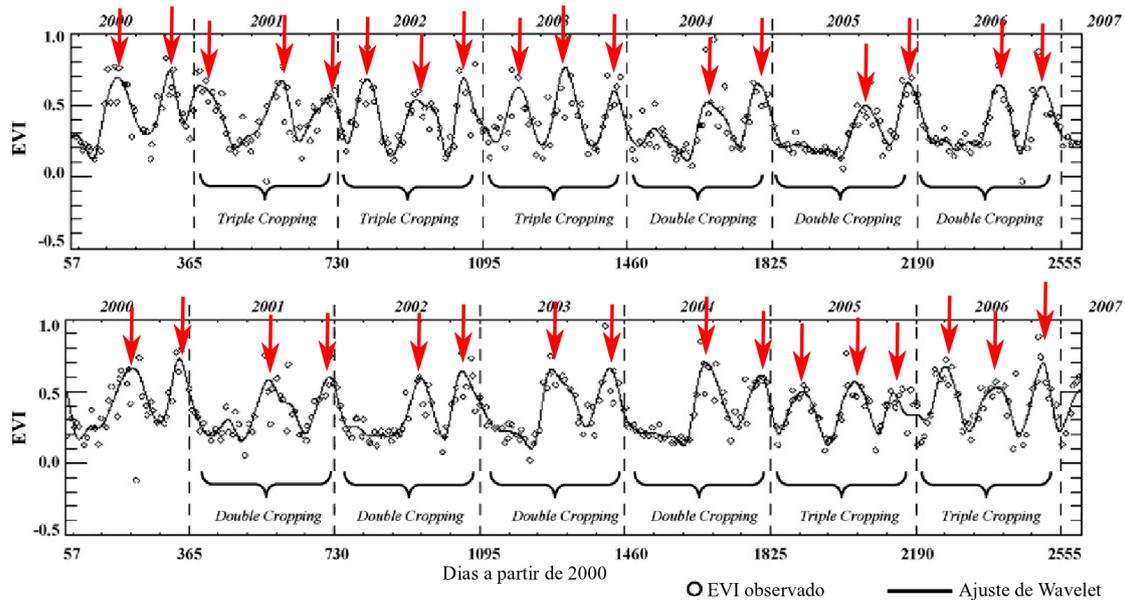


Figura 5.2 - Intensificação no cultivo de arroz. As setas em vermelho indicam os picos de sazonalidade.

Fonte: Adaptado de Sakamoto et al. (2009).

É importante notar que, em muitas culturas, o início do crescimento da biomassa tende a ser seguido por um período de rápido crescimento de área foliar, que se mantém estável até o ponto do decaimento da atividade fenológica para níveis próximos a zero. Nesse contexto, Zhang et al. (2003) afirmam que existem quatro datas chave para a caracterização das fases fenológicas da vegetação em escala anual: verdejamento, maturidade, senescência e dormência (Ver Figura 5.3). Em linha semelhante, Jönsson e Eklundh (2004) desenvolveram uma ferramenta (TIMESAT) para análise de séries temporais de dados de satélite com base em métricas extraídas das sazonalidades. A Figura 5.3 mostra a extração de pontos relevantes da série, sendo (a) o início e (b) o fim da sazonalidade, (c) e (d) os níveis de 80%, (e) o ponto de maior valor, (f) a amplitude, (g) o comprimento, e (h) e (i) as integrais mostrando o efeito cumulativo do crescimento da vegetação durante a estação. Assim como as métricas gerados por Zhang et al. (2003), as informações produzidas pelo TIMESAT também podem ser utilizadas para monitorar e detectar mudanças de uso do solo e cobertura da terra.

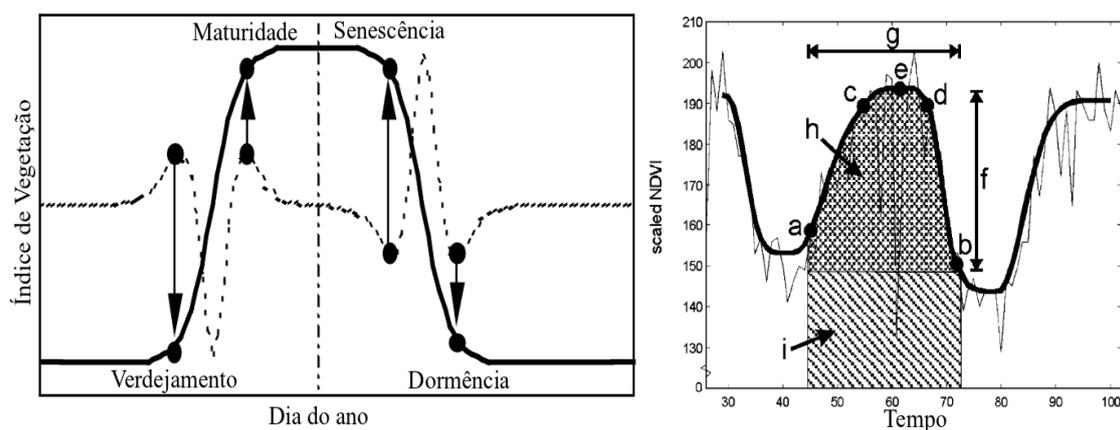


Figura 5.3 - Representação gráfica das métricas de caracterização fenológica; no lado esquerdo datas chave para caracterização das sazonalidades. Na direita, métricas de caracterização da sazonalidade extraídas com a ferramenta TIMESAT.

Fonte: Adaptado de Jönsson e Eklundh (2004).

5.1 Classificação inicial dos dados

Como visto, os dados de sensoriamento remoto com alta resolução temporal possuem um enorme potencial para o monitoramento e detecção das mudanças de uso e cobertura da terra por meio da análise das características sazonais. Entretanto, é essencial o conhecimento dos tipos de sistemas de cultivo normalmente adotadas na região de interesse, bem como seu calendário agrícola. Além disso, algum conhecimento histórico, cultural, econômico e político da região de interesse, pode ajudar na compreensão das trajetórias de uso, já que o uso da terra é determinado por uma mistura de fatores relativos a diversas áreas do conhecimento.

A falta de estudos sobre dados temporais de índice de vegetação relacionados a tipos de uso do solo e cobertura da terra foi um grande problema enfrentado no presente trabalho, já que não foi encontrado material suficiente que fizesse atribuição de padrões temporais de índice de vegetação para classes de uso e cobertura do solo. Este é mais um fato que justifica o desenvolvimento de um sistema computacional capaz de fazer a classificação automática desses padrões. Entretanto, não tendo um especialista disponível para classificar precisamente uma quantidade significativa de amostras, e nem mesmo tendo a verdade terrestre, os padrões utilizados neste trabalho foram pré-classificados por meio de inspeção visual. Nessa inspeção, optou-se por investigar quatro grandes classes: floresta, desflorestamento, pastagem e agricultura.

Foram utilizadas séries temporais obtidas a partir do trabalho desenvolvido por Freitas (2012), disponíveis no site <https://www.dsr.inpe.br/laf/series/>. A área geográfica do globo onde essas séries foram tomadas é um retângulo composto por cerca de $10,5 \text{ km}^2$ no leste do estado de Mato Grosso, Brasil, cujas coordenadas geográficas opostas estão situadas nos pontos $(-12,331945, -52,422560)$ e $(-12,372355, -52,458480)$. Dentro desta área de estudo foi selecionada uma amostra representativa de 168 píxeis, onde cada píxel corresponde a um quadrado com 250 m de lado. A Figura 5.4 ilustra a localização de onde as séries foram selecionadas.

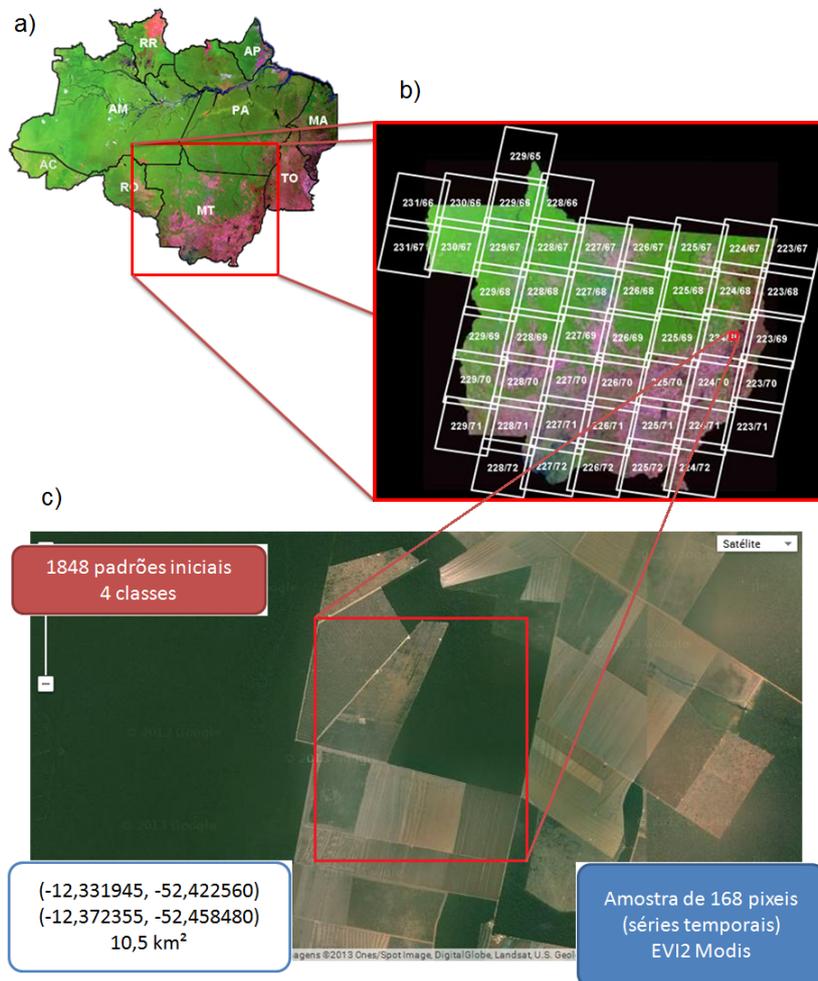


Figura 5.4 - Localização da área de estudo abordada.

Cada píxel possui uma série temporal de índice de vegetação associada a ele. As séries temporais correspondem a 11 anos de observações do índice de vegetação EVI2, de agosto de 2000 a junho de 2011. Considerando o período total de 11 anos,

a informação disponível é constituída de 265 observações por píxel. As séries foram divididas por ano – que é constituída de 23 observações – e cada ano tomado como um padrão de uso e cobertura da terra. Sendo assim o número total de padrões temporais de uso do solo e cobertura da terra para a área seleccionada foi de 1848 (168×11). Pelo fato de um padrão ter exatamente 23 (cerca de $365/16$) observações por ano por píxel, que produz um total de 253 (23×11) observações por píxel, a diferença entre o total de observações de uma série temporal seleccionada e o conjunto de padrões que essa série produz teve que ser descartada. Portanto, foram desconsideradas as 12 primeiras das 265 observações de cada série original (ver Figura 5.5 como ilustração).

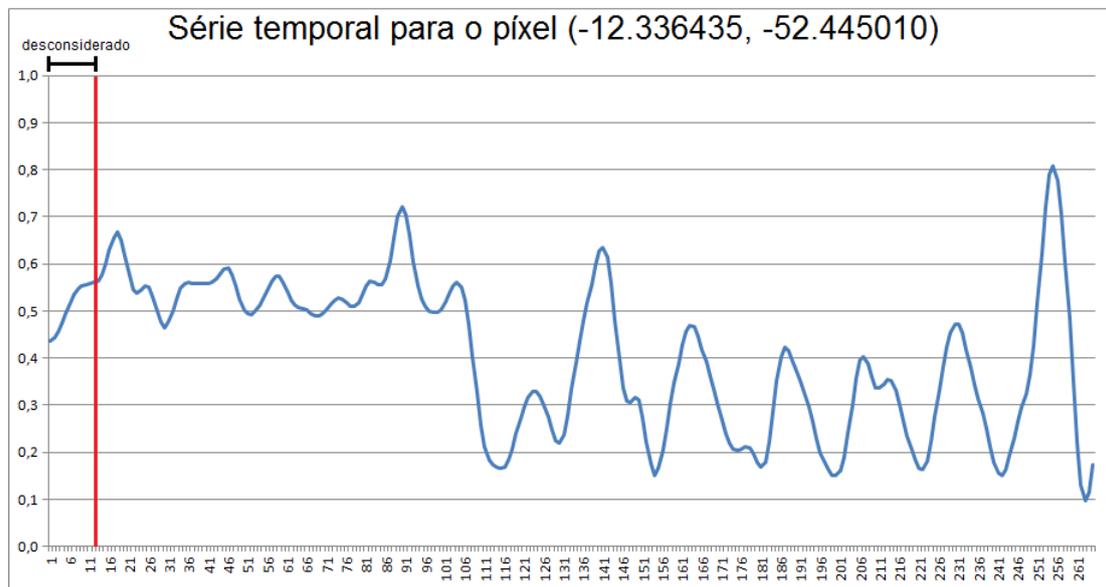


Figura 5.5 - Ilustração das observações desconsideradas.

Uma vez que os padrões foram escolhidos e classificados por meio de inspeção visual, realizou-se um procedimento de filtragem dos dados a fim de diminuir a quantidade de possíveis padrões classificados erroneamente. Mais detalhes sobre esse procedimento pode ser visto na seção 6.2.

6 MÉTODOS E RESULTADOS

Para cumprir o objetivo proposto neste trabalho, foi investigado se a utilização do sistema neuro-difuso ANFIS seria o mais indicado para tratar o problema em comparação com a utilização da rede neural MLP. Os resultados preliminares deste trabalho, e que resultaram no estudo e desenvolvimento do MLP difuso, estão descritos na Seção 6.1.

Definido o método principal para solucionar o problema, foi necessário selecionar um conjunto de séries temporais de índice de vegetação e subdividir essas séries em padrões temporais de uso do solo e cobertura da terra, como descrito na Seção 5.1. Após isso, a fim de diminuir possíveis erros da classificação inicial um processo de filtragem dos dados foi realizado e descrito na Seção 6.2. Com as séries escolhidas, os padrões selecionados, e o sistema neural difuso desenvolvido, os parâmetros livres do sistema foram definidos de acordo com os métodos expostos na Seção 6.3 e na Seção 6.4.

Foi então realizado uma série de testes para determinar qual sistema, dentre quatro opções especificadas na Seção 6.5, era capaz de melhor classificar padrões temporais de uso do solo e cobertura da terra utilizando séries temporais de índice de vegetação. Determinou-se o melhor método para esse objetivo e, então, utilizando esse método, foi realizado um estudo paramétrico (ver Seção 6.6) para verificar quais seriam os melhores valores para os parâmetros do método selecionado. Finalmente, uma validação do sistema comparando os resultados obtidos por ele e a classificação do TerraClass 2010, numa área selecionada do leste do estado brasileiro do Mato Grosso. Essa validação pode ser vista na Seção 6.7.

6.1 Testes preliminares

A princípio, era de nosso interesse utilizar o sistema de inferência adaptativo neuro-difuso ANFIS para cumprir com o objetivo proposto. Contudo, esse sistema não se mostrou adequado para o problema, pelo motivo básico de necessitar de um período muito grande de treinamento.

O ANFIS se mostrou inadequado devido a sua explosão combinatória em relação as regras. Suponha que n é o número de variáveis de entrada e p é o número de partições difusas para cada variável. Desse modo, o número de regras é p^n . Para o caso de um padrão anual de uso do solo e cobertura da terra ter $n = 23$ observações, e se for usado $p = 3$ – como o ANFIS indica para configuração padrão – o número

de regras seria de 3^{21} , ou seja, 10,4 bilhões de regras. Para $p = 2$, então, o número de regras cairia para 2,1 milhões. Mesmo assim, o ANFIS não conseguiu criar a rede sendo executado num computador com a seguinte configuração:

- Processador Intel Core i5 2,53 GHz;
- Memória principal (RAM) de 4 Gb;
- Sistema Operacional Windows 7 de 64 bits.

Além disso, como essa é apenas a estrutura por onde as entradas serão propagadas e retropropagadas a cada padrão, e depois a cada época, o processamento era muito alto, pois o sistema teria mais de $3 \cdot p^n$ nós, e cada nó possuindo *loops* de complexidade $O(n^2)$.

Para contornar o problema, o número de observações do padrão anual foi diminuído para 12, depois para 8 e por fim para 6. Isso, calculando a média aritmética a cada duas, três e quatro observações, respectivamente. A Figura 6.1 ilustra a degradação de um padrão de 23 observações (pontos) para padrões de 12, 8 e 6 pontos.

23 Pontos	12 Pontos	8 Pontos	6 Pontos
0,5043	0,5047	0,5097	0,5180
0,5051			
0,5196	0,5314		
0,5431		0,5625	
0,5672	0,5722		0,5611
0,5771			
0,5639	0,5500	0,5348	
0,5360			
0,5044	0,4948		0,4977
0,4851		0,4954	
0,4904	0,5006		
0,5108			
0,5368	0,5481	0,5548	0,5588
0,5593			
0,5684	0,5696		
0,5708		0,5706	
0,5718	0,5705		0,5664
0,5692			
0,5658	0,5623	0,5571	
0,5587			
0,5467	0,5420		0,5393
0,5373		0,5356	
0,5338	0,5338		

Figura 6.1 - Exemplo de degradação de um padrão temporal de uso do solo e cobertura da terra.

Em resumo, alguns testes foram realizados utilizando um conjunto classificado com

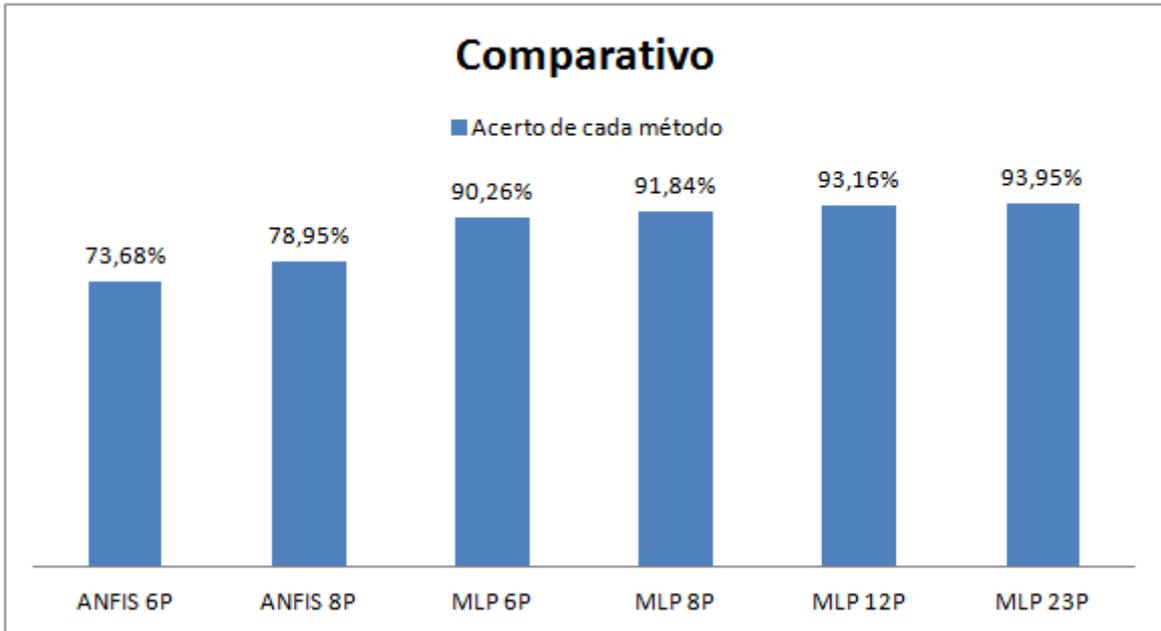


Figura 6.2 - Comparativo preliminar entre métodos.

76 padrões temporais de uso do solo e cobertura da terra, sendo que metade foi utilizado para treinamento e a outra metade para validação. Esses padrões foram obtidos em exemplos do trabalho de Freitas (2012). O mesmo conjunto foi testado com 12, 8 e 6 pontos em cada padrão. Os resultados com o sistema ANFIS não pareciam satisfatórios, então os mesmos testes foram repetidos com a utilização da rede neural MLP clássica. Os resultados estão ilustrados na Figura 6.2.

Foi visto que a diminuição do número de pontos, dentro do padrão, era prejudicial ao sistema e que o ANFIS obteve resultados insatisfatórios em comparação ao MLP em todos os testes. Além disso, o ANFIS só conseguiu construir o conjunto de regras utilizando padrões com 6 e 8 pontos, mesmo assim o tempo de treinamento era muito alto.

Por esses motivos, o uso do ANFIS para tratar do problema proposto foi descartado.

6.2 Filtragem dos dados

Sendo o padrão uma subsequência de uma série temporal do índice de vegetação EVI2, que foi representada como um vetor de valores reais contendo 23 elementos, foram calculados, para cada elemento do vetor, a média e o desvio padrão das classes do conjunto inicial, como ilustrado na Figura 6.3, isso com o intuito de diminuir a dispersão do conjunto inicial de dados classificados. Foram selecionados apenas os

padrões totalmente contidos dentro da faixa definida pelo valor médio mais ou menos k desvios-padrão ($k \cdot \sigma$). A Figura 6.4 ilustra este processo de filtragem para $k = 1$.

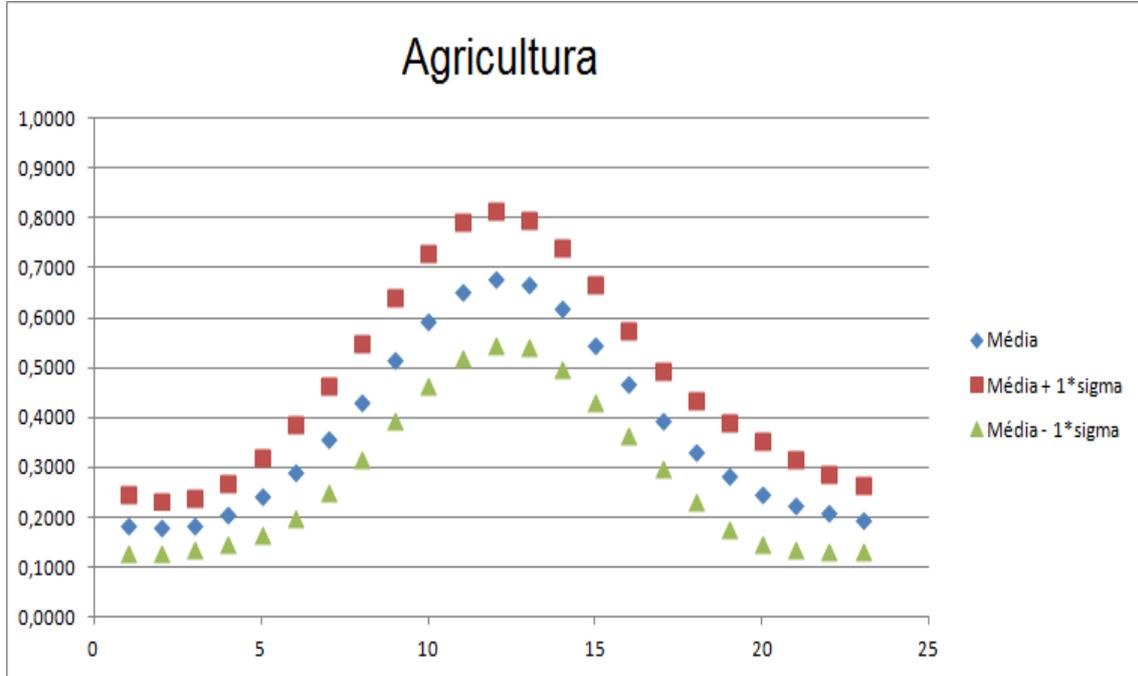


Figura 6.3 - Classe Agricultura com a indicação dos valores de média + ou - $k = 1$ desvios-padrão ($k \times \sigma$).

O desvio padrão – Equação 6.1 – mede a dispersão dos valores individuais em torno da média – Equação 6.2.

$$\sigma(x) = \sum_{i=1}^N \sqrt{\frac{(\bar{x} - x_i)^2}{N}}, \quad (6.1)$$

onde

$$\bar{x} = \sum_{i=1}^N \frac{x_i}{N}. \quad (6.2)$$

Para essa seleção de padrões, uma variabilidade menor implicaria em uma especialização maior, entretanto uma grande especialização prejudicaria na generalização, que é crucial para um sistema de classificação como a rede MLP. Dessa forma, é muito importante encontrar um valor para k que diminua a variabilidade de pa-

drões existentes em cada classe, eliminando erros da classificação inicial, e, ao mesmo tempo, preserve a existência de padrões diversificados da mesma classe.

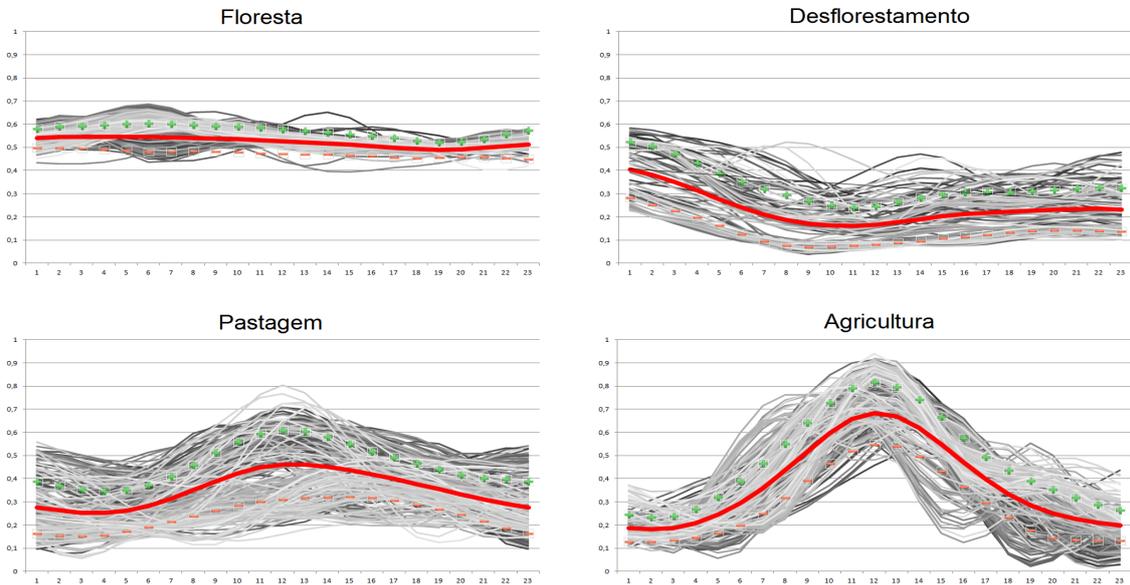


Figura 6.4 - Classes de séries temporais de índice de vegetação com indicação dos vetores média e um desvio padrão.

Diversos subconjuntos filtrados foram gerados variando o valor de k de 0,5 até 3,0, com passo de 0,5. Para testar se o subconjunto resultante da filtragem era satisfatório usou-se uma validação cruzada *4-folder* para a rede MLP tradicional. Por meio desses testes, chegou-se a conclusão que o melhor valor para k era 1 (um). Para valores de $k < 1$ a quantidade de padrões para algumas classes era quase que zerada e para $k > 2$ a quantidade de padrões era praticamente constante, sendo assim, a fim de comparação, foi decidido que seriam criados dois subconjuntos de dados um para $k = 1$ e um para $k = 2$.

O conjunto I foi formado de 205 padrões obtidos usando $k = 1$, e o conjunto II de 1363 padrões obtidos usando $k = 2$. Os valores de $k = 1$ e $k = 2$ correspondem a atitudes mais ou menos conservadoras, respectivamente, considerando-se uma ausência de conhecimentos sólidos e/ou verdade do solo. O conjunto I contém 139 padrões de floresta, 18 de desflorestamento, 35 de pastagem e 13 da agricultura. O conjunto II contém 641 padrões de floresta, 100 de desflorestamento, 279 de pastagem e 343 da agricultura.

6.3 Aspectos práticos do treinamento de redes MLP

Um dos problemas enfrentados no treinamento de redes neurais do tipo MLP, com treinamento por retropropagação, diz respeito à definição de seus parâmetros. A seleção dos parâmetros de treinamento do algoritmo é um processo pouco compreendido e pequenas diferenças nestes parâmetros podem levar a grandes diferenças tanto no tempo de treinamento como na generalização obtida.

Existem inúmeras pesquisas no sentido de encontrar uma fórmula ideal que determine a configuração ótima de uma rede neural para um dado problema. No entanto, o que se tem hoje são apenas sugestões que estão fundamentadas em experiências vividas por especialistas no assunto e no bom senso.

Deve-se ter em mente que é preciso obter um modelo que não seja muito rígido a ponto de não modelar fielmente os dados, mas que também não seja excessivamente flexível a ponto de modelar também o ruído presente neles. A ideia é que a rede responda de acordo com as características presentes nos dados de entrada e não exatamente igual aos dados de entrada.

São apresentados, a seguir, os principais parâmetros a serem ajustados numa rede perceptron multicamadas e os métodos mais usados para definí-los.

- **Número de camadas ocultas e de neurônios em cada camada oculta**

Dois dos aspectos mais importantes a serem considerados na configuração de uma rede perceptron multicamadas é o número de camadas ocultas, juntamente com o número de neurônios em cada uma dessas camadas. Usualmente, as quantidades de neurônios nas camadas de entrada e saída são definidas de acordo com o padrão de treinamento, sendo a quantidade de neurônios na camada de entrada igual a quantidade de variáveis no padrão de entrada e a quantidade de neurônios na camada de saída igual ao número de elementos no vetor de saída do padrão. Já nas camadas escondidas, a escolha do número de camadas e de neurônios por camada pode ser determinante para o sucesso da utilização da rede. A determinação do número ótimo de camadas e de neurônios em cada camada oculta é feita de modo empírico. Várias redes são criadas e treinadas, ao final deve ser selecionada aquela com o mínimo de camadas e neurônios que fornece o menor erro médio quadrático de forma rápida e eficiente.

Deve-se ter cuidado para não utilizar nem unidades demais, o que pode levar a rede a memorizar os dados de treinamento (*overfitting*), ao invés de extrair as características gerais que permitirão a generalização, nem um número muito pequeno, que pode forçar a rede a gastar tempo em excesso tentando encontrar uma representação ótima.

Hecht-Nielsen (1989) adaptou um teorema relativo à aproximação de funções, desenvolvido pelo matemático russo Kolmogorov em 1957, às redes neurais artificiais multicamadas, reformulando o teorema da seguinte maneira: “Dada uma função contínua arbitrária $f : [0, 1]^n \rightarrow R^m$, $f(x) = y$, existe sempre, para f , uma implementação exata com uma rede neural artificial de três camadas, sendo a camada de entrada um vetor de dimensão n , a camada oculta composta por $(2n + 1)$ neurônios, e a camada de saída com m neurônios representando as m componentes do vetor y ” (KOVACS, 2006).

Para problemas de classificação uma única camada oculta é suficiente para realizar a classificação. Já os problemas de predição, apesar de funcionarem bem com uma camada oculta, podem ter sua capacidade de predição melhorada com uma segunda camada oculta, sem comprometer a capacidade de generalização da rede. Por outro lado, o uso de uma terceira camada oculta não melhora a capacidade da rede e requer muito mais tempo para treinar.

No presente estudo, foi percebido que uma camada escondida seria suficiente para tratar o problema, já o número de neurônios nessa camada foi testado variando de 6 a 50, sendo que quanto maior o número de neurônios nessa camada maior o tempo despendido para o processamento do treinamento. O número de neurônios escolhido foi de dez, já que apresentou um desempenho satisfatório e o aumento da quantidade de neurônios diminuiu a taxa de acerto do sistema.

- **Normalização dos dados de entrada e saída**

O processo de normalização é crítico para o resultado do treinamento. Se os valores de entrada e saída não forem da mesma magnitude, os pesos deverão ser ajustados fortemente para compensar estas diferenças, o que não é feito de forma eficiente nos algoritmos de treinamento como a retropropagação. Além disso, as funções de ativação típicas, como a função logística e a função tangente hiperbólica, não podem distinguir entre valores de magnitude muito grande (veja o gráfico destas funções na Figura

6.5).

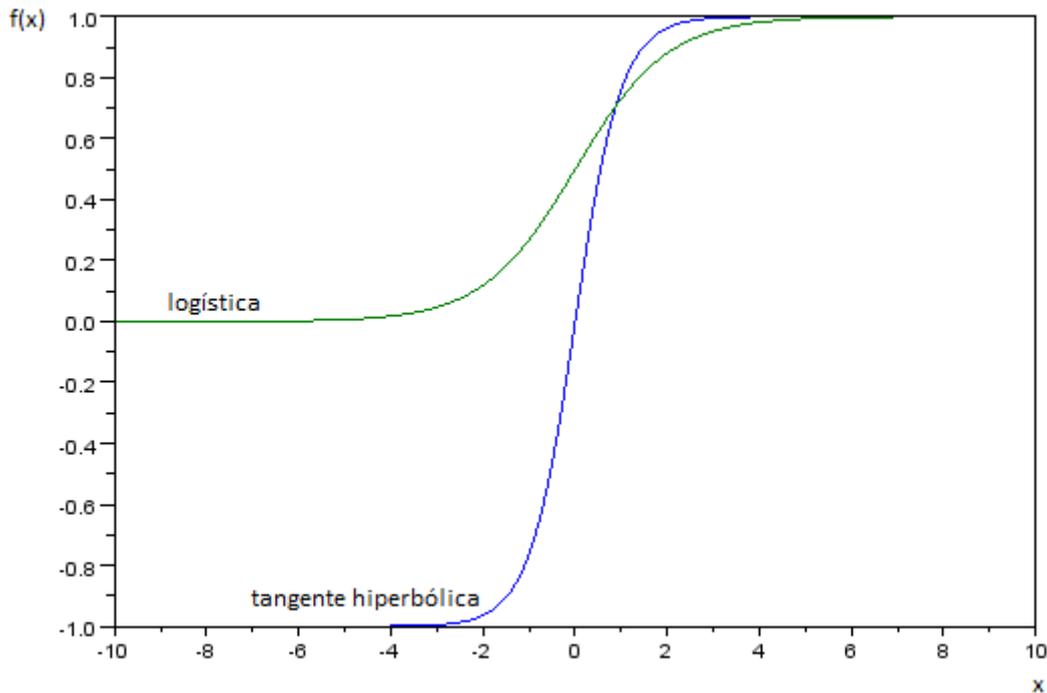


Figura 6.5 - Duas das funções de ativação que podem ser usadas no MLP.

Para evitar estes problemas, os valores de entrada e saída podem ser normalizados utilizando, entre outros, um dos seguintes métodos:

- Normalizar para o intervalo $[0, 1]$, dividindo pelo valor maior:

$$x_i^N = \frac{x_i}{x_i^{max}}. \quad (6.3)$$

Esse método garante que o maior valor de entrada será 1, mas não garante que 0 seja o menor. Uma limitante deste método é que ele não aproveita integralmente o intervalo de aplicação da função de ativação.

- Normalizar para o intervalo $[0, 1]$, garantindo que $x^{min} = 0$ e $x^{max} = 1$:

$$x_i^N = \frac{x_i - x_i^{min}}{x_i^{max} - x_i^{min}}. \quad (6.4)$$

Este método é significativamente melhor que o anterior, ele expande a

intervalo de definição da variável utilizando todo o alcance da função de ativação. Além disso, todas as variáveis do conjunto de dados terão a mesma distribuição, o que melhora a eficiência do treinamento.

- Normalizar para o intervalo $[-1, 1]$, convertendo o valor médio a 0. Este método é chamado de normalização de média zero e é obtido pela fórmula:

$$x_i^N = \frac{x_i - \bar{x}_i}{\max[x_i^{\max} - \bar{x}_i, \bar{x}_i - x_i^{\min}]}, \quad (6.5)$$

onde x_{med} é o valor médio da variável no conjunto de dados.

Este método, como o anterior, utiliza todo o alcance da função de ativação e garante que todas as variáveis tenham a mesma distribuição.

Aqui, como o valor do índice de vegetação EVI2 varia entre 0 e 1, não foi necessário utilizar um método de normalização. Além disso, foram utilizadas as séries temporais disponibilizadas filtradas com a transformada *wavelet*, já que possuíam variação mais suave entre as observações em comparação com os dados brutos.

- **Ajuste da taxa de aprendizado e do coeficiente *momentum***

A taxa de aprendizado e o coeficiente *momentum* são dois importantes parâmetros que controlam a efetividade do treinamento da rede com o algoritmo de retropropagação. A taxa de aprendizado é uma constante positiva que controla a velocidade com que os fatores de peso são ajustados baseado no valor da correção do gradiente descendente. O coeficiente *momentum* é um peso extra que acelera a velocidade de ajuste dos fatores de peso. Ele permite que o ajuste dos pesos saia de mínimos locais utilizando um valor alto no começo do treinamento que tende a zero na medida que a rede neural é treinada.

O parâmetro taxa de aprendizado tem grande influência durante o processo de treinamento da rede neural. Uma taxa de aprendizado muito baixa torna o aprendizado lento, ao passo que uma taxa de aprendizado muito alta provoca oscilações no treinamento e impede a convergência do processo de aprendizado. Geralmente seu valor varia de 0,1 a 1,0.

O coeficiente *momentum* pode ou não ser utilizado durante o treinamento e seu valor varia de 0,0 (não utilizado) a 1,0.

No caso presente, tanto a taxa de aprendizado quanto o coeficiente *momentum* foram ajustados e testados de um valor inicial até um valor final.

Sendo que o valor mínimo utilizado para o *momentum* foi 0,0, diferente da taxa de aprendizado que teve como valor mínimo 0,2. A taxa de aprendizado foi atribuída com os valores 0,2, 0,4, 0,6, 0,8 e 1,0. Já o coeficiente *momentum* recebeu os valores 0,0, 0,3, 0,5, 0,7 e 0,9. Após os testes tomou-se 0,6 para o valor padrão da taxa de aprendizado e 0,5 para o *momentum*.

- **Escolha da função de ativação**

A função de ativação para um neurônio artificial do perceptron multicamadas deve representar se o neurônio está ativo ou inativo (“ligado” ou “desligado”), assim ela deve ser uma função contínua, monotônica crescente e limitada. Além disso, devido ao fato da derivada da função de ativação ser utilizada no cálculo de atualização dos pesos sinápticos, ela deve ser diferenciável em todo o seu domínio.

As duas funções de ativação mais utilizadas são:

- Logística;
- Tangente hiperbólica.

Uma vantagem da função de ativação tangente hiperbólica em relação a função logística é que o alcance da tangente hiperbólica (intervalo da imagem) é naturalmente maior $\|[-1, 1]\| = 2$ do que a da função logística $\|[0, 1]\| = 1$, ou seja, apresenta uma resposta maior para desvios pequenos nos valores de entrada.

A tangente hiperbólica apresenta valores tanto positivos quanto negativos para os valores de entrada e resposta enquanto a função logística apresenta resposta positiva para todos os valores de entrada. Esta característica, combinada com a normalização de média zero permite dar sentido aos valores de saída do neurônio: 0 representa o estado normal do neurônio enquanto ± 1 representa valores de resposta muito baixos/altos.

As duas funções foram testadas e para o problema proposto a função de ativação logística apresentou desempenho superior.

- **Critério de parada**

Existem vários métodos para a determinação do momento em que o treinamento de uma rede neural deve ser encerrado. Uma boa determinação destes critérios é fundamental para um bom treinamento e conseqüentemente uma boa generalização. Os critérios de parada mais utilizados são:

- Número de ciclos: define o número de ciclos de treinamento, ou seja, o número de vezes em que o conjunto de treinamento é apresentado à rede. Um número excessivo de ciclos pode levar a rede à perda do poder de generalização (*overfitting*). Por outro lado, com um pequeno número de ciclos a rede pode não chegar ao seu melhor desempenho (*underfitting*).
- Erro: consiste em encerrar o treinamento após o erro médio quadrático ficar abaixo de um valor pré-definido. Vale lembrar que um erro médio quadrático muito pequeno não implica necessariamente numa boa generalização (i.e. bom desempenho com dados não vistos anteriormente). Este valor depende muito do problema.
- Número de ciclos ou erro: pode-se também utilizar uma combinação entre um número de ciclos máximo e um valor mínimo para o erro. Desta forma, o treinamento é encerrado quando qualquer um desses critérios é atendido.
- Validação (*Best Model*): na técnica de parada pela validação, o treinamento é interrompido periodicamente a cada vez que um número determinado de ciclos é atingido, assim é realizada uma estimação de erro da rede sobre o conjunto de dados de teste. A partir do momento em que o erro medido no conjunto de teste apresentar crescimento, o treinamento é encerrado. O que se deseja com esta técnica é descobrir o momento exato em que a rede começa a perder generalização.

O critério de parada utilizado foi o *número de ciclos ou erro*, ou seja, se o número de ciclos atingir um valor máximo pré-estabelecido ou o erro atingir um valor mínimo desejado o treinamento é encerrado. Cada um desses parâmetros recebeu dois valores, sendo o erro mínimo de 0,0001 e 0,00001 e o máximo de ciclos de 1000 e 10000. O menor valor para o erro e o maior para o número de ciclos ocasionaram o já mencionado *overfitting* ao sistema, assim os valores complementares (0,0001 e 1000, respectivamente) foram os escolhidos.

6.4 Definição dos parâmetros do sistema

Todos os parâmetros livres do sistema, como taxa de aprendizado, coeficiente *momentum*, número de camadas escondidas, número de neurônios em cada camada escondida, coeficiente do grau de ambiguidade e outros, foram obtidos por meio da realização de várias simulações.

O primeiro parâmetro a ser definido foi o número de camadas escondidas, obedecendo ao proposto por Hecht-Nielsen (1989), após isso, foi o número de neurônios nessa camada, usando como parâmetros de decisão a acurácia do sistema – para uma validação cruzada *4-folder* sendo executada dez vezes – e o tempo de processamento gasto nos treinamentos. Nesses testes não foi utilizado o coeficiente *momentum* e a taxa de aprendizado foi definida como 0,5. O critério de parada foi o número de iterações (ciclos) de 1000. Definidos esses valores, foi a vez de selecionar conjuntamente a taxa de aprendizado e coeficiente *momentum*. Após isso, função de ativação e o critério de parada. Todos de acordo com o descrito na seção 6.3.

Os dois parâmetros difusos do cálculo da saída difusa desejada foram definidos como se parecia pertinente, recebendo valores de 0,5 a 2,0 com intervalo de 0,5. Tanto f_d quanto f_e tomaram o valor de 2,0 como o mais adequado.

Já o parâmetro m do cálculo do grau de ambiguidade recebeu todos os valores no intervalo real $[0, 2]$ com passo de 0,2. Sendo que para valores menores do que 1 o grau de ambiguidade está tendo maior influência na atualização dos pesos da rede e para valores maiores do que 1 tem menor influência. O valor 0 (zero) foi utilizado para testar a não influência do coeficiente grau de ambiguidade no cálculo de atualização dos pesos da rede. O valor definido para m foi 0,2

6.5 Comparação entre métodos e entre conjuntos de dados

Uma vez obtidos os conjuntos de treinamento e validação, foi realizada uma série de testes nos conjuntos, através de um processo de validação cruzada *4-folder* (sendo três partes para treinamento e uma parte para validação). Foram aplicados quatro métodos sobre os dados: o perceptron multicamadas tradicional (MLP), o MLP utilizando o grau de ambiguidade (MLP-G), o MLP usando a saída difusa desejada (MLP-D) e o MLP utilizando grau de ambiguidade e saída difusa desejada (MLP-GD). Os métodos MLP-D e MLP-GD obtiveram os melhores resultados, com $f_d = f_e = 2$ em uma série de experimentos de tentativa e erro. A Tabela 6.1 apresenta os resultados para os conjuntos I e II, mostrando a médias de acerto para

cada método usado na validação 4-*folder* considerando as classes individuais, e os resultados gerais.

Tabela 6.1 - Taxa de acerto individual para as classes F (floresta), D (desflorestamento), P (pastagem) e A (agricultura), usando os conjuntos I e II, e acerto global dos conjuntos I e II, para os métodos MLP tradicional (MLP), MLP utilizando grau de ambiguidade (MLP-G), MLP usando saída difusa desejada (MLP-D) e MLP utilizando ambos, grau de ambiguidade e saída difusa desejada (MLP-GD).

Sistema	Acurácia									
	Individual								Global	
	I				II				I	II
	F	D	P	A	F	D	P	A		
MLP	100.00%	94.74%	94.63%	67.52%	98.33%	98.20%	86.40%	89.73%	89.22%	93.16%
MLP-G	100.00%	93.91%	96.57%	83.33%	98.95%	97.47%	88.12%	89.45%	93.45%	93.50%
MLP-D	100.00%	95.29%	95.10%	79.62%	99.83%	97.67%	89.83%	80.86%	92.50%	92.05%
MLP-GD	100.00%	94.58%	96.84%	92.78%	99.47%	97.80%	86.86%	90.08%	96.05%	93.55%

Observou-se que todos os métodos produziram resultados muito bons, em geral, para os dois conjuntos I e II. O conjunto I produziu melhores resultados do que o conjunto II para acurácia individual das classes floresta (F) e pastagem (P), ao passo que o conjunto II produziu melhores resultados para desflorestamento (D). Em relação a agricultura, o conjunto II foi melhor do que o I, exceto para o método MLP-GD, que é também o melhor resultado para esta classe. No que se refere a precisões individuais, o método MLP-GD é ou o melhor do que suas contrapartes para todas as classes em ambos os conjuntos I e II, ou com resultados muito semelhantes às melhores. No que diz respeito a resultados globais, o método MLP-GD saiu melhor para ambos os conjuntos I e II. Para os resultados globais, o conjunto I se saiu melhor do que o conjunto II para os métodos MLP-GD e MLP-D e pior para os métodos MLP e MLP-G. Os melhores resultados globais foram produzidos pelo método MLP-GD para o conjunto I. Em conclusão, o método MLP-GD e o conjunto I foram os mais apropriados para o tratamento dos dados.

Usando as forças de disparo y dos neurônios da camada de saída da rede, obteve-se um índice de confiança c do sistema de classificação, mediante a aplicação da entropia de Shannon (PINEDA, 2006), calculada por meio da Equação 6.6.

$$c = 100 \cdot \left(1 - \frac{-\sum y'_i(p_0) \cdot \log y'_i(p_0)}{\log |C|}\right), \quad (6.6)$$

onde p_0 é o padrão de entrada sendo avaliado, y'_i é a força de disparo normalizada

do neurônio da camada de saída de índice i e $|C|$ o número neurônios nessa camada.

Este índice é ilustrado na Figura 6.5, que indica os resultados obtidos para um conjunto de píxeis selecionados, para ambos os conjuntos, I (lado esquerdo) e II (lado direito). Os resultados descritos são os obtidos pelo método MLP-GD, ou seja, usando tanto saída difusa desejada e cálculo do grau de ambiguidade. Na figura, foi indicada a evolução do índice de vegetação no interior de cada um dos 11 padrões anuais. Também mostra-se a classificação de cada padrão anual usando tanto um código de cores quanto letras: F, D, P e A. A diferença na tonalidade em cada padrão indica a confiança do sistema no seu próprio julgamento, dado pela entropia dentro do padrão. Quanto mais escura a cor, menor a entropia e maior a confiança.

O primeiro píxel, representado em a) e b) na figura, mostra uma série uniforme de padrões florestais. Pode-se ver que a diferença entre os conjuntos I e II é a certeza do sistema em sua classificação: o conjunto I produz resultados com menor entropia e, portanto, com maior certeza. A mesma tendência pode ser vista nos outros píxeis na figura.

O segundo píxel, representado em c) e d) na figura, indica que, de acordo com o conjunto I, a área era floresta nos dois primeiros anos, depois sofreu desflorestamento no terceiro ano, em seguida, tornou-se pastagem no quarto ano, e por fim, foi utilizada para a agricultura nos restantes sete anos, seguindo uma ordem lógica. No conjunto II, o segundo ano foi classificado como pastagem, antes de se tornar desflorestamento no terceiro ano e novamente pasto no quarto ano, o que não é consistente em termos de evolução temporal dos padrões da região.

O terceiro píxel ilustra um resultado menos satisfatório. É visto que, usando o conjunto I, após a pastagem começar no 6º ano, o uso torna-se floresta no 10º ano para voltar à pastagem no 11º ano. No ano 10, a região foi, provavelmente, também coberta por pastagem, embora, possivelmente, um pasto mais sujo do que os outros anos. Note, porém, que o sistema dá baixa certeza na avaliação. Usando o conjunto II, apenas uma sequência de padrões florestais, seguido por uma sequência de padrões de pastagem foram detectados, não apresentando a etapa esperada de desflorestamento entre eles.

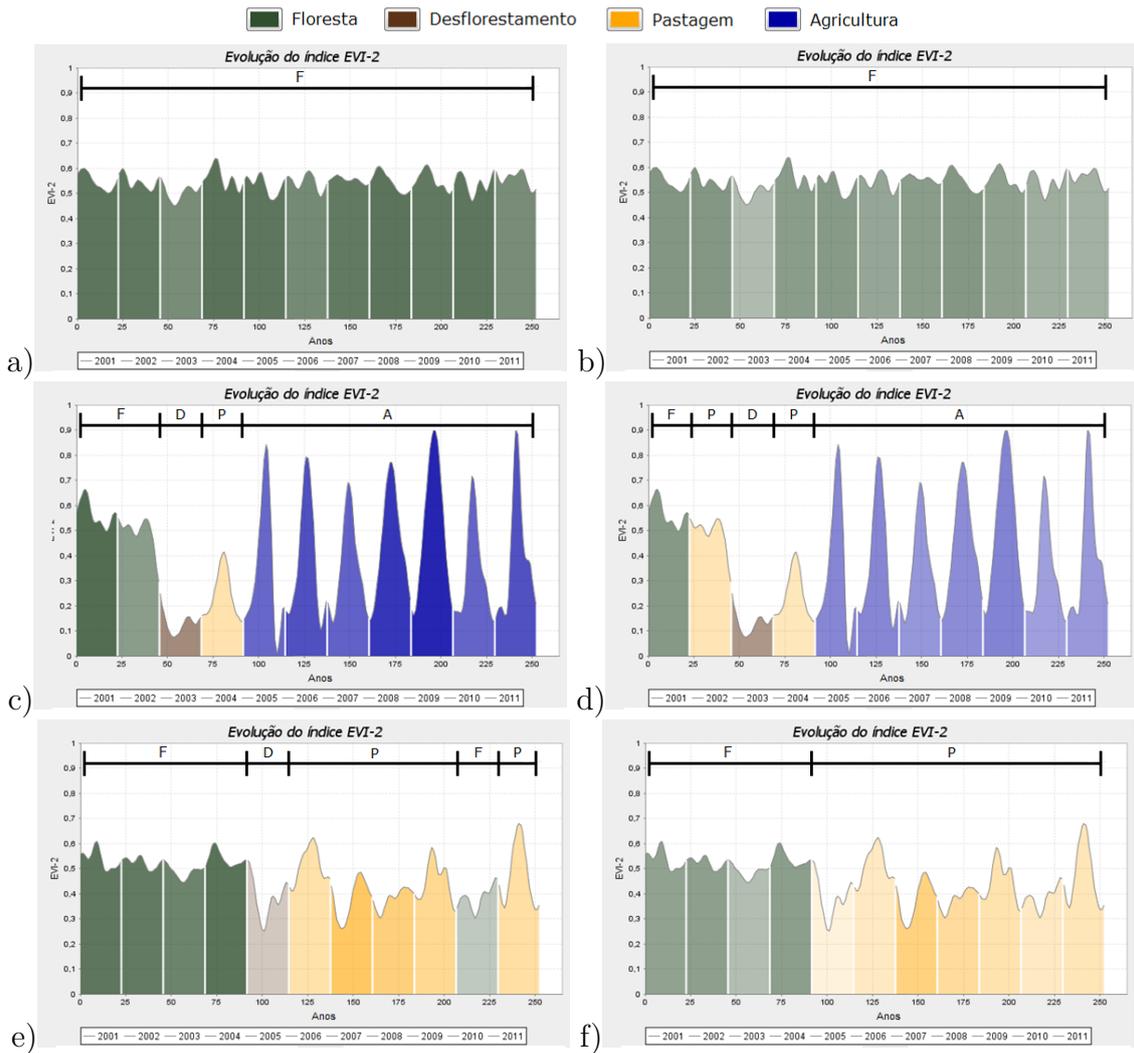


Figura 6.6 - Resultados dos experimentos em 3 pixels, utilizando os conjuntos I (lado esquerdo) e II (lado direito).

6.6 Estudo paramétrico

Definido o método MLP-GD como o melhor dentre os quatro analisados, foi realizado, então, um estudo paramétrico para determinar quais eram os valores mais adequados para os parâmetros do sistema: taxa de aprendizado – que variou de 0,1 a 1,0, com passo de 0,1 –, coeficiente *momentum* – que variou de 0,0 a 1,0, com passo de 0,2 –, coeficiente *m* do grau de ambiguidade – que variou de 0,0 a 2,0, com passo de 0,2 – e o número de neurônios na camada intermediária – que variou de 6 a 50, com passo de 2 –. Formou-se então 15.180 combinações de valores para esses parâmetros. Ou seja, mais de 15 mil configurações do sistema.

O método MLP-GD foi executado para cada uma dessas configurações, numa vali-

dação 4-*folder*, utilizando o conjunto I. Em 4,81% das 15.180 configurações, o uso de 28 neurônios na camada intermediária obteve as maiores taxas de acerto. Entretanto, a tabela 6.2 mostra que a variação do número de neurônios na camada intermediária quase que não interfere na acurácia do sistema para valores acima de 16 neurônios.

Tabela 6.2 - Percentual de vezes que cada quantidade de neurônios na camada escondida obteve as melhores taxas de acerto.

Número de neurônios	Melhores resultados	Número de neurônios	Melhores resultados
6	1,58%	30	4,57%
8	3,13%	32	4,61%
10	3,65%	34	4,54%
12	4,11%	36	4,56%
14	4,43%	38	4,57%
16	4,61%	40	4,57%
18	4,61%	42	4,54%
20	4,64%	44	4,59%
22	4,70%	46	4,57%
24	4,73%	48	4,59%
26	4,61%	50	4,66%
28	4,81%		

Para se visualizar as taxas de acerto das demais configurações, criou-se tabelas no estilo “mapas de calor” (ver Figuras 6.6).

Os resultados mostrados nas Figuras 6.6, referem-se às taxas de acerto global do método MLP-GD, utilizando o conjunto I. Cada figura diz respeito a variação das taxas de aprendizado, de 0,1 a 1,0, as linhas são as variações do coeficiente *momentum*, de 0,0 a 1,0, e cada coluna é referente a um valor do parâmetro difuso *m* do grau de ambiguidade, de 0,0 a 2,0.

É possível perceber que quanto maior o coeficiente *momentum* melhores são os resultados. Entretanto, há um limite para ele, o valor 0,8, pois assim que esse coeficiente atinge o valor 1,0 o método só é capaz de classificar corretamente 25% dos padrões – apenas os padrões de floresta. Já, o coeficiente *m* do grau de ambiguidade aumenta a acurácia do método conforme seu valor é decrementado. Por outro lado, as Figuras 6.6b, 6.6c e 6.6d mostram que o valor 0,2 trás melhores resultados do que o valor 0,0.

6.7 Comparação com classificações TerraClass 2010

Após os treinamentos e testes usando os conjuntos de dados I e II, buscou-se um meio de validar e comparar os resultados do sistema proposto com dados reais ou, no mínimo, mais próximos da realidade. Pensou-se então, assim como em Maus (2013), na utilização dos dados do projeto TerraClass para essa tarefa.

O projeto TerraClass tem por objetivo classificar o uso do solo e cobertura da terra das áreas desflorestadas de toda Amazônia Legal (ver Figura 6.8a). Ele é baseado nas áreas desflorestadas mapeadas e publicadas pelo Projeto PRODES (Monitoramento da Floresta Amazônica Brasileira por Satélite) e como resultado disponibiliza, por meio de sítio eletrônico, o mapeamento do uso do solo e cobertura da terra para essa região. Segundo o seu site oficial [http : //www.inpe.br/cra/projetos_pesquisas/terraclass2010.php](http://www.inpe.br/cra/projetos_pesquisas/terraclass2010.php), a execução do projeto se deu através da mobilização de equipes do INPE (Instituto Nacional de Pesquisas Espaciais) e da Embrapa (Empresa Brasileira de Pesquisa Agropecuária).

O TerraClass possui dois produtos, TerraClass 2008 e TerraClass 2010, e possibilita a comparação da evolução do uso do solo e cobertura da terra da área amazônica entre esses dois mapeamentos.

Com esses dados, foi possível fazer uma avaliação da acurácia dos resultados obtidos pelo sistema neural difuso proposto por este trabalho. No entanto, uma consideração importante que deve ser feita sobre esta comparação é que o TerraClass apresenta a classificação de uma observação pontual no tempo, ou seja, a classificação TerraClass não está relacionada a um padrão temporal de observações, como no caso dos padrões aceitos no sistema proposto.

Para esta tarefa, selecionou-se uma área de 8,325 mil hectares, entre os pontos geográficos opostos $(-12, 397050, -52, 478685)$ e $(-12, 318475, -52, 397865)$. Veja a Figura 6.8b com a seleção da área das séries temporais obtidas pelo sítio eletrônico [http : //www.dsr.inpe.br/la.f/series/](http://www.dsr.inpe.br/la.f/series/) disponível através do trabalho de Freitas (2012).

A mesma área obtida por meio do mapa disponibilizado pelo **TerraClass 2010** pode ser vista na Figura 6.9.

Tem-se a área quadriculada, pois foi preciso que os dados do TerraClass 2010 tivessem a mesma resolução espacial que os produtos MODIS, já que este é o tamanho do píxel das séries temporais usadas no treinamento. O píxel é um quadrado de 250

metros de lado.

A imagem foi degradada para esta resolução, sendo que cada píxel foi obtido calculando-se a classe dominante presente nele. Assim, obteve-se a Figura 6.10a. Tem-se, então, um conjunto de 1332 píxeis, sendo 36 linhas e 37 colunas.

Em contraposição, a classificação da área selecionada, para a subsérie temporal 2009-2010, obtida pelo sistema neural difuso com o conjunto de treinamento I pode ser vista na Figura 6.10b.

Deve-se levar em consideração que o número de classes observadas no Projeto TerraClass é muito maior do que o que foi usado neste trabalho. Especificamente para a área selecionada, foi possível ver sete classes no mapa TerraClass 2010, sendo que os conjunto de treinamento do sistema usado possuía apenas quatro classes.

Considerando que as classes “Floresta”, “Agricultura” e “Desflorestamento”, que estão presentes nas duas classificações – TerraClass 2010 e sistema neural difuso – correspondem à mesma classe e que “Regeneração com pasto” e “Pasto limpo” do TerraClass, correspondem a classe “Pastagem” do sistema neural difuso, foram classificados de maneira correta 81% dos píxeis analisados, sendo que o sistema acertou 95% de floresta, 41% de pastagem e 78% de agricultura, por outro lado não foi possível perceber a presença de desflorestamento em nenhuma das duas classificações. Para esses números, não foram considerados os píxeis das classes “Vegetação secundária” e “Área não observada”.

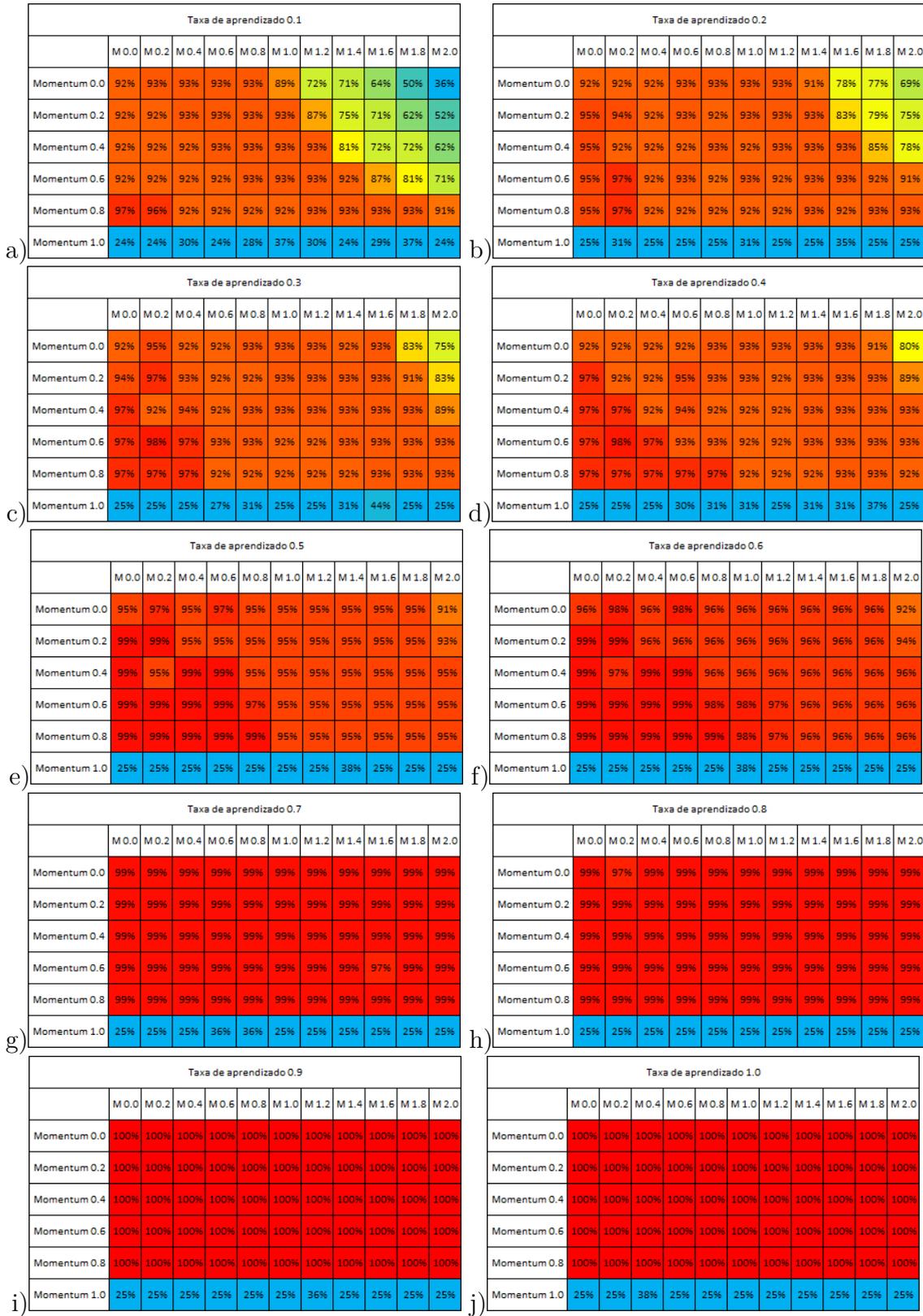
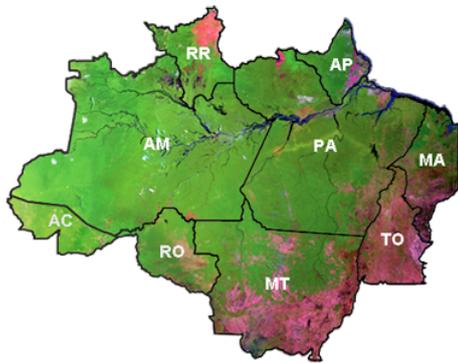


Figura 6.7 - Mapas de calor das taxas de acerto.



a) Amazônia legal.



b) Área de estudo.

Figura 6.8 - Amazônia legal e área de estudo.

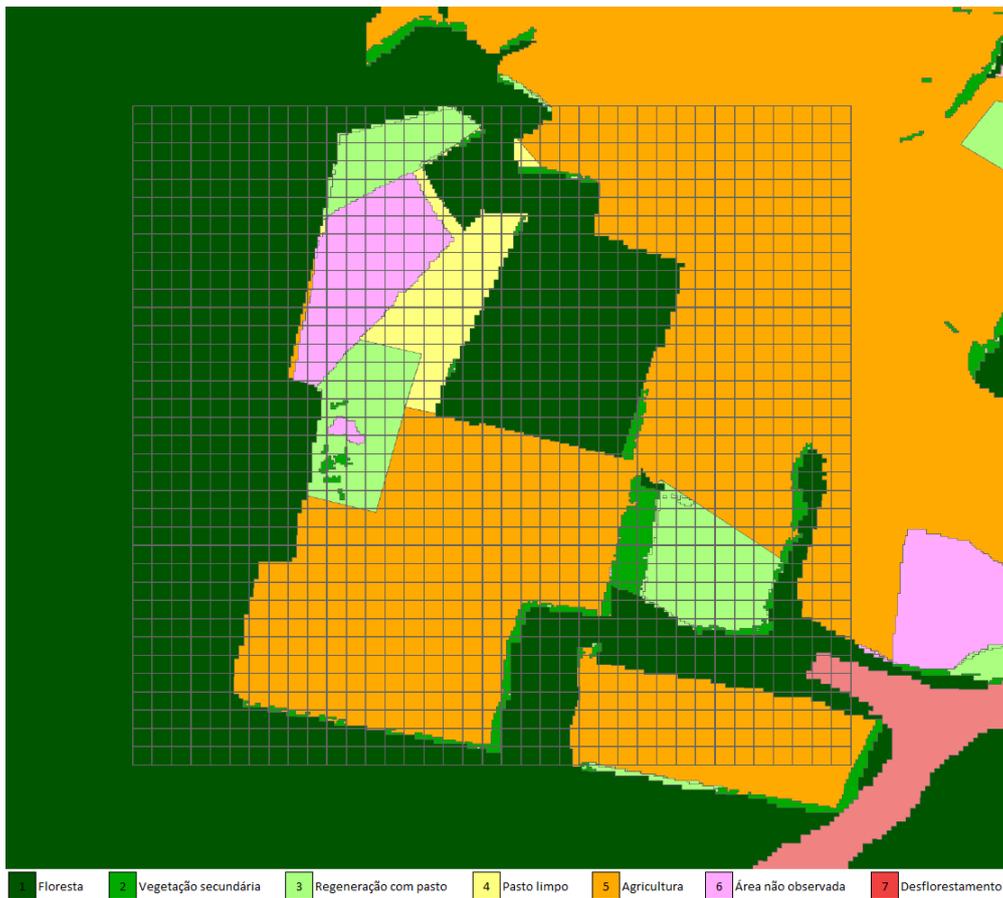


Figura 6.9 - Mapeamento TerraClass 2010.

Fonte: Maus (2013).



Figura 6.10 - Comparação entre classificações. a) Classificação TerraClass 2010 b) Classificação do Sistema Neural Difuso.

7 CONCLUSÕES

A análise contínua das mudanças de cobertura da terra e dos impactos causados em outros processos do sistema terrestre dependem da consistência e aumento na disponibilidade de séries temporais com alta frequência. Além disso, é essencial um melhor entendimento das relações entre medições físicas e processos biofísicos (LAMBIN; LINDERMAN, 2006). Neste sentido, o uso de produtos e técnicas de sensoriamento remoto e georreferenciamento em análise ambiental tem se tornado cada vez mais comum. No caso do uso do solo e cobertura da terra, essas técnicas contribuem de forma expressiva para a velocidade, eficiência e confiabilidade nas análises que envolvem os processos de degradação da vegetação natural, monitoramento dos recursos florestais, desenvolvimento de políticas de conservação, assim como vários outros fatores que podem causar mudanças na vegetação.

Investigou-se, neste trabalho, a combinação de sistemas difusos e redes neurais para a classificação de padrões de uso do solo e cobertura da terra na Amazônia brasileira, utilizando-se séries temporais de índices de vegetação a partir do sensor MODIS da NASA. Para isso, usou-se, além do *Perceptron* Multicamadas (MLP) tradicional, três implementações difusas: MLP com grau de ambiguidade, MLP com saída difusa desejada e MLP com grau de ambiguidade e saída difusa desejada. Esses métodos foram aplicados a uma área de estudo de aproximadamente $10,5\text{km}^2$ no leste do estado de Mato Grosso, na Amazônia brasileira. No geral, todas as implementações testadas parecem promissoras, entretanto a combinação do grau de ambiguidade e da saída difusa desejada forneceu os melhores resultados. A aplicação de um processo de filtragem para os dados brutos provou-se útil na ausência de um especialista para identificar a verdade terrestre a partir de dados de sensores remotos. No entanto, esta técnica também possui suas limitações ilustradas pela dificuldade do sistema em distinguir entre floresta e pasto sujo em alguns casos.

Portanto, o uso de sistemas neuro-difusos para classificação do uso e cobertura da terra, a partir de séries temporais de índices de vegetação, é justificável pela sua alta taxa de acerto nos testes e apesar da rede neural artificial MLP clássica ter obtido bons resultados, as modificações difusas trouxeram ainda mais benefícios à rede.

Para validação do sistema, foi realizada a comparação entre a classificação calculada pelo sistema neuro-difuso e a classificação do mapeamento TerraClass 2010 de uma região de 8,325 mil hectares no leste do estado brasileiro do Mato Grosso. Esta comparação foi positiva, pois o sistema classificou de maneira correta 81% dos píxeis analisados. É importante notar que mesmo havendo diferença entre a observação

pontual no tempo, do Projeto TerraClass, e os padrões de observações temporais, do sistema proposto, a taxa de acerto obtida pode ser considerada muito boa e um incentivo para novos estudos.

É possível ver que a maior quantidade de erros do sistema produzido está nos píxeis de intercessão entre duas coberturas, normalmente agricultura e floresta. Uma justificativa plausível para esse fato é que nesses píxeis estejam presentes mais do que uma cobertura e as quantidades de verde da floresta e da agricultura se misturem, formando um padrão mais parecido com a pastagem.

Deseja-se, no futuro, expandir a área geográfica de estudos, selecionando mais regiões do país e com variadas classes de uso do solo e cobertura da terra. Isso com o intuito de investigar se é possível criar um sistema capaz de diferenciar as diversas características fenológicas das coberturas vegetais presentes no Brasil ou se, para isso, seria necessário criar um sistema especializado para cada região do país. Ademais, é pensado o desenvolvimento de um aplicativo móvel – para *smartphones* e *tablets* – capaz de disponibilizar ao grande público a funcionalidade de visualização de séries temporais de índice de vegetação já classificadas, ano a ano, com suas mais prováveis classes de uso do solo e cobertura da terra. Este seria um aplicativo valioso para gerentes e empresários dos ramos de agricultura e pecuária, para pesquisadores científicos e agentes ambientais e também para agricultores familiares. Por fim, busca-se, dia a dia, formas de melhorar os resultados e a precisão do sistema para que sua confiabilidade seja frequentemente incrementada.

Como frutos desta pesquisa, além do presente trabalho de dissertação de Mestrado em Computação Aplicada pelo Instituto Nacional de Pesquisas Espaciais, dois artigos foram publicados e apresentados: *Uso de Perceptrons Multicamadas Difusos para a Classificação de Padroes em Séries Temporais*, aceito no X Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2013) (PIMENTEL et al., 2013b); e *Using Fuzzy Multilayer Perceptrons for the Classification of Time Series*, aceito no Tenth International Workshop on Fuzzy Logic and Applications (WILF 2013) (PIMENTEL et al., 2013a).

REFERÊNCIAS BIBLIOGRÁFICAS

- AACH, J.; CHURCH, G. M. Aligning gene expression time series with time warping algorithms. **Bioinformatics**, Oxford University Press, v. 17, n. 6, p. 495–508, 2001. 8
- ADAMI, M.; FREITAS, R. M. d.; PADOVANI, C. R.; SHIMABUKURO, Y. E.; MOREIRA, M. A. Estudo da dinâmica espaço-temporal do bioma pantanal por meio de imagens modis. **Pesquisa agropecuária brasileira**, SciELO Brasil, v. 43, n. 10, p. 1371–1378, 2008. 8
- AGUIAR, A. P. D. d. **Modelagem de mudança do uso da terra na Amazônia**: explorando a heterogeneidade intrarregional. 206 p. Tese (Doutorado em Sensoriamento Remoto) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2012. 1
- AITKENHEAD, M. J.; MCDONALD, A. J. S. A neural network face recognition system. **Engineering applications of artificial intelligence**, Elsevier, v. 16, n. 3, p. 167–176, 2003. 10
- BARBOSA, H. A. **Análise espaço temporal de índice de vegetação AVHRR/NOAA e precipitação na região nordeste do Brasil em 1982-85**. 170 p. (INPE-7237-TDI/690). Dissertação (Mestrado em Sensoriamento Remoto) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 1998. 2
- BRADLEY, P. S.; FAYYAD, U. M.; REINA, C. et al. Scaling clustering algorithms to large databases. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 4., 1998, New York, NY, USA. **Proceedings...** New York: ACM, 1998. p. 9–15. ISBN 1-57735-070-7. 12
- BRIASSOULIS, H. Analysis of land use change: theoretical and modeling approaches. **Regional Research Institute**, West Virginia University, v. 1, n. 1, 2000. 1
- CÂMARA, G.; VALERIANO, D. M.; SOARES, J. V. **Metodologia para o cálculo da taxa anual de desmatamento na Amazônia Legal**. São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2006. Relatório técnico. Disponível em: <<http://www.obt.inpe.br/prodes/metodologia.pdf>>. 5

CANUTO, A. M. d. P. **Combining neural networks and fuzzy logic for applications in character recognition**. 196 p. Thesis (Doctor of Philosophy) — University of Kent at Canterbury, Canterbury, 2001. 15, 27, 38, 39, 41

CHATFIELD, C. **The analysis of time series: an introduction**. 6. ed. Florida: CRC press, 2013. 326 p. 8

COLLISTER, A. A.; LAHAV, O. Anz: estimating photometric redshifts using artificial neural networks. **Publications of the astronomical society of the Pacific**, JSTOR, v. 116, n. 818, p. 345–351, 2004. 10

DEDE, G.; HÜSNÜ, S. M. Speech recognition with artificial neural networks. **Digital signal processing**, Elsevier, v. 20, n. 3, p. 763–768, 2010. 10

DEFRIES, R. Terrestrial vegetation in the coupled human-earth system: contributions of remote sensing. **Environment and resources**, v. 33, p. 369–390, 2008. 2

DEFRIES, R. S.; HOUGHTON, R. A.; HANSEN, M. C.; FIELD, C. B.; SKOLE, D.; TOWNSHEND, J. Carbon emissions from tropical deforestation and regrowth based on satellite observations for the 1980s and 1990s. **National academy of sciences**, National Acad Sciences, v. 99, n. 22, p. 14256–14261, 2002. 2

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. et al. Maximum likelihood from incomplete data via the em algorithm. **Journal of the royal statistical society**, v. 39, n. 1, p. 1–38, 1977. 12

DESAI, A. A. Gujarati handwritten numeral optical character reorganization through neural network. **Pattern recognition**, Elsevier, v. 43, n. 7, p. 2582–2589, 2010. 10

DIMLA, D. E.; LISTER, P. M. On-line metal cutting tool condition monitoring: ii: tool-state classification using multi-layer perceptron neural networks.

International journal of machine tools and manufacture, Elsevier, v. 40, n. 5, p. 739–768, 2000. 31

DING, H.; TRAJCEVSKI, G.; SCHEUERMANN, P.; WANG, X.; KEOGH, E. Querying and mining of time series data: experimental comparison of representations and distance measures. **VLDB endowment**, v. 1, n. 2, p. 1542–1552, 2008. 10

DUBOIS, D.; PRADE, H. Possibility theory. **Computational complexity: theory, techniques, and applications**, Springer, New York, v. 1, p. 2240–2252, 2012. 22

DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern classification**. 2. ed. New Jersey: John Wiley & Sons, 2012. 680 p. 9, 10

EADS, D. R.; HILL, D.; DAVIS, S.; PERKINS, S. J.; MA, J.; PORTER, R. B.; THEILER, J. P. Genetic algorithms and support vector machines for time series classification. **International symposium on optical science and technology**, International society for optics and photonics, v. 1, p. 74–85, 2002. 11

EPIPHANIO, J. C. N.; GLERIANI, J. M.; FORMAGGIO, A. R.; RUDORFF, B. F. T. Índices de vegetação no sensoriamento remoto da cultura do feijão. **Pesquisa agropecuária brasileira**, v. 31, p. 445–454. 6

FORMIGONI, M. d. H.; XAVIER, A. C.; LIMA, J. S. D. S. Análise temporal da vegetação na região do nordeste através de dados evi do modis. **Ciência florestal**, v. 21, n. 1, p. 1–8, 2011. 8

FREEMAN, S. **Biological science**. 2. ed. New Jersey: Pearson Prentice Hall, 2005. 1392 p. 16

FREITAS, R. M. d. **Laboratório virtual para visualização e caracterização do uso e cobertura da terra utilizando imagens de sensoriamento remoto**. 235 p. (INPE-17003-TDI/1882). Tese (Doutorado em Computação Aplicada) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2012. 2, 43, 44, 47, 51, 65

GALFORD, G. L.; MUSTARD, J. F.; MELILLO, J.; GENDRIN, A.; CERRI, C. C.; CERRI, C. E. P. Wavelet analysis of modis time series to detect expansion and intensification of row-crop agriculture in brazil. **Remote Sensing of Environment**, Elsevier, v. 112, n. 2, p. 576–587, 2008. 43, 44

GEURTS, P. Pattern extraction for time series classification. In: **PRINCIPLES OF DATA MINING AND KNOWLEDGE DISCOVERY**, 5., 2001, Freiburg, Germany. **Proceedings...** Freiburg: Springer, 2001. p. 115–127. ISBN 978-3-540-42534-2. 10

GOMIDE, F.; GUDWIN, R.; TANSCHKEIT, R. Conceitos fundamentais da teoria de conjuntos fuzzy, lógica fuzzy e aplicações. In: **INTERNATIONAL FUZZY SYSTEMS ASSOCIATION WORLD CONGRESS**, 6., 1995, Sao Paulo, Brazil. **Anais...** São Paulo: IFSA, 1995. p. 1–38. 22

GOWARD, S. N.; MARKHAM, B.; DYE, D. G.; DULANEY, W.; YANG, J. Normalized difference vegetation index measurements from the advanced very high resolution radiometer. **Remote sensing of environment**, Elsevier, v. 35, n. 2, p. 257–277, 1991. 7

GÜLER, E.; SANKUR, B.; KAHYA, Y. P.; RAUDYS, S. Visual classification of medical data using mlp mapping. **Computers in biology and medicine**, Elsevier, v. 28, n. 3, p. 275–287, 1998. 31

HAN, J.; KAMBER, M.; PEI, J. **Data mining: concepts and techniques**. 3. ed. Waltham, Massachusetts: Morgan kaufmann, 2006. 709 p. 9, 10

HAYES, D. J.; SADER, S. A. Comparison of change-detection techniques for monitoring tropical forest clearing and vegetation regrowth in a time series. **Photogrammetric engineering and remote sensing**, v. 67, n. 9, p. 1067–1075, 2001. 6

HAYKIN, S. S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001. 900 p. 10, 15, 17, 26, 32, 33, 34, 37, 38

HECHT-NIELSEN, R. Theory of the backpropagation neural network. In: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 1989, Washington, DC, USA. **Proceedings...** Washington: IEEE, 1989. p. 593–605. 55, 60

HWANG, R. C.; HUANG, H. C.; CHEN, Y. J.; HSICH, J. G.; CHAO, H. Adaptive power signal prediction by non-fixed neural network model with modified fuzzy back-propagation learning algorithm. In: INFORMATION, COMMUNICATIONS AND SIGNAL PROCESSING, 1., 1997, Singapore. **Proceedings...** Singapore: IEEE, 1997. p. 689–692. ISBN 0-7803-3676-3. 37

INDRO, D. C.; JIANG, C. X.; PATUWO, B. E.; ZHANG, G. P. Predicting mutual fund performance using artificial neural networks. **Omega**, Elsevier, v. 27, n. 3, p. 373–380, 1999. 31

JAMES, M. E.; KALLURI, S. N. V. The pathfinder avhrr land data set: an improved coarse resolution data set for terrestrial monitoring. **International Journal of Remote Sensing**, Taylor & Francis, v. 15, n. 17, p. 3347–3363, 1994.

7

JANG, J. S. R. Anfis: Adaptive-network-based fuzzy inference system. **Systems, man and cybernetics, IEEE transactions on, IEEE**, v. 23, n. 3, p. 665–685, 1993. 27, 28

JEONG, J. H.; KIM, H.; KIM, D. S.; LEE, S. Y. Speaker adaptation based on judge neural networks for real world implementations of voice-command systems. **Information sciences**, Elsevier, v. 123, n. 1, p. 13–24, 2000. 31

JIN, Y. Fuzzy modeling of high-dimensional systems: complexity reduction and interpretability improvement. **Fuzzy Systems, IEEE transactions on, IEEE**, v. 8, n. 2, p. 212–221, 2000. 15

JÖNSSON, P.; EKLUNDH, L. Timesat: A program for analyzing time-series of satellite sensor data. **Computers & Geosciences**, Elsevier, v. 30, n. 8, p. 833–845, 2004. 45, 46

JUSTICE, C. O.; VERMOTE, E.; TOWNSHEND, J. R. G.; DEFRIES, R.; ROY, D. P.; HALL, D. K.; SALOMONSON, V. V.; PRIVETTE, J. L.; RIGGS, G.; STRAHLER, A. et al. The moderate resolution imaging spectroradiometer (modis): land remote sensing for global change research. **Geoscience and Remote Sensing, IEEE Transactions on, IEEE**, v. 36, n. 4, p. 1228–1249, 1998. 7

KALPAKIS, K.; GADA, D.; PUTTAGUNTA, V. Distance measures for effective clustering of arima time-series. In: INTERNATIONAL CONFERENCE ON DATA MINING, 1., 2001, San Jose, CA, USA. **Proceedings...** San Jose: IEEE, 2001. p. 273–280. ISBN 0-7695-1119-8. 8

KAMPOURAKI, A.; MANIS, G.; NIKOU, C. Heartbeat time series classification with support vector machines. **Information technology in biomedicine, IEEE transactions on, IEEE**, v. 13, n. 4, p. 512–518, 2009. 11

KELLER, J. M.; HUNT, D. J. Incorporating fuzzy membership functions into the perceptron algorithm. **Pattern analysis and machine intelligence, IEEE transactions on, IEEE**, v. 7, p. 693–699, 1985. 31, 37, 41

KEOGH, E. J.; PAZZANI, M. J. An enhanced representation of time series which allows fast and accurate classification, clustering and relevance feedback. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 4., 1998, New York, NY, USA. **Proceedings...** New York: ACM, 1998. p. 239–243. ISBN 1-57735-070-7. 8, 11

KOVACS, Z. L. **Redes neurais artificiais: fundamentos e aplicações**. 4. ed. São Paulo: Livraria da Física, 2006. 55

LACRUZ, M. P.; SANTOS, J. R.; CONFORTE, J. C. Utilização de séries temporais de imagens modis/terra e análise harmônica para o monitoramento de corredores ecológicos. **Simpósio brasileiro de sensoriamento remoto (SBSR)**, v. 12, p. 3223–3228, 2005. 8

LACRUZ, M. S. P.; JÚNIOR, M. d. A. S. Uso de séries temporais evi/modis e análise harmônica para o estudo da bacia do rio taquari. **Revista brasileira de cartografia**, v. 59, n. 1, p. 9–15, 2011. 2

LAMBIN, E. F.; LINDERMAN, M. Time series of remote sensing data for land change science. **Geoscience and remote sensing, IEEE transactions on, IEEE**, v. 44, n. 7, p. 1926–1928, 2006. 71

LAWRENCE, C. E.; REILLY, A. A. An expectation maximization (em) algorithm for the identification and characterization of common sites in unaligned biopolymer sequences. **Proteins: Structure, Function, and Bioinformatics**, Wiley Online Library, v. 7, n. 1, p. 41–51, 1990. 12

LEE, C.-C. Fuzzy logic in control systems: fuzzy logic controller. **Systems, man and cybernetics, IEEE transactions on, IEEE**, v. 20, n. 2, p. 419–435, 1990. 24

LENT, R. **Cem bilhões de neurônios: conceitos fundamentais de neurociência**. 1. ed. Rio de Janeiro: Atheneu, 2004. 698 p. 15

LIN, C.-T.; LEE, C. S. **Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems**. 1. ed. New Jersey: Prentice-Hall, Inc., 1996. 797 p. 17, 29

LIN, J.; VLACHOS, M.; KEOGH, E.; GUNOPULOS, D. Iterative incremental clustering of time series. In: INTERNATIONAL CONFERENCE ON EXTENDING DATABASE TECHNOLOGY, 9., 2004, Heraklion, Crete, Greece. **Proceedings...** Heraklion: Springer, 2004. p. 106–122. ISBN 978-3-540-21200-3. 12

LIN, J.; WILLIAMSON, S.; BORNE, K.; DEBARR, D. Pattern recognition in time series. **Advances in Machine Learning and Data Mining for Astronomy**, Citeseer, v. 1, p. 617–645, 2012. 5, 8, 9, 10, 12

LIPPMANN, R. P. An introduction to computing with neural nets. **ASSP Magazine, IEEE transactions on**, v. 4, n. 2, p. 4–22, 1987. 10

- LLOYD, D. A phenological classification of terrestrial vegetation cover using shortwave vegetation index imagery. **Remote sensing**, Taylor & Francis, v. 11, n. 12, p. 2269–2279, 1990. 7
- MACQUEEN, J. Some methods for classification and analysis of multivariate observations. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 14., 1967, Berkeley, CA, USA. **Proceedings...** Berkeley: University of California Press, 1967. p. 281–297. 12
- MAMDANI, E. H. Advances in the linguistic synthesis of fuzzy controllers. **International journal of man-machine studies**, Elsevier, v. 8, n. 6, p. 669–678, 1976. 24
- MAUS, V. W. **Detecção e monitoramento de mudanças de uso da terra com dados de sensoriamento remoto de alta resolução temporal**. 26 p. Exame de Qualificação (Doutorado em Ciência do Sistema Terrestre) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2013. 6, 65, 68
- MEHROTRA, K.; MOHAN, C. K.; RANKA, S. **Elements of artificial neural networks**. 2. ed. Cambridge: Massachusetts Institute of Technology, 1996. 344 p. 15, 26, 31, 37
- MÜLLER, B.; REINHARDT, J.; STRICKLAND, M. T. **Neural networks: an introduction**. 2. ed. New York: Springer Verlag, 2013. 331 p. ISBN 3-540-60207-0. 15
- MYNENI, R. B.; KEELING, C. D.; TUCKER, C. J.; ASRAR, G.; NEMANI, R. R. Increased plant growth in the northern high latitudes from 1981 to 1991. **Nature**, v. 386, n. 6626, p. 698–702, 1997. 7
- NASSAR, A. M.; RUDORFF, B. F. T.; ANTONIAZZI, L. B.; AGUIAR, D. A. d.; BACCHI, M. R. P.; ADAMI, M. et al. Prospects of the sugarcane expansion in brazil: impacts on direct and indirect land use changes. **Sugarcane ethanol: contribution to climate change mitigation and the environment**, Wageningen Academic Publishers, v. 1, p. 63–93, 2008. 7
- PAL, S. K.; MITRA, S. Multilayer perceptron, fuzzy sets, and classification. **Neural networks, IEEE transactions on**, IEEE, v. 3, n. 5, p. 683–697, 1992. 31, 32, 36, 37, 38, 39, 41
- PIMENTEL, T.; RAMOS, F. M.; SANDRI, S. Using fuzzy multilayer perceptrons for the classification of time series. In: INTERNATIONAL WORKSHOP ON

FUZZY LOGIC AND APPLICATIONS, 10., 2013, Genoa, Italy. **Proceedings...** Genoa: Springer, 2013. p. 60–67. ISBN 978-3-319-03199-6. 72

_____. Uso de perceptrons multicamadas difusos para a classificação de padrões em séries temporais. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL, 10., 2013, Fortaleza, CE, Brasil. **Proceedings...** Fortaleza: ENIAC 2013, 2013. p. 12. 72

PINEDA, J. O. d. C. **A entropia segundo Claude Shannon**: o desenvolvimento do conceito fundamental da teoria da informação. 118 p. Dissertação (Mestrado em História da Ciência) — Universidade Católica de São Paulo, São Paulo, 2006. 61

PUTTER, T. D.; LOUTRE, M.-F.; WANSARD, G. Decadal periodicities of Nile river historical discharge (ad 622–1470) and climatic implications. **Geophysical research letters**, Wiley Online Library, v. 25, n. 16, p. 3193–3196, 1998. 5

REZENDE, S. O. **Sistemas inteligentes**: fundamentos e aplicações. 1. ed. Barueri, SP: Editora Manole, 2003. 527 p. ISBN 85-204-1683-7. 24

RODRIGUES, P. P.; GAMA, J.; PEDROSO, J. P. Hierarchical clustering of time-series data streams. **Knowledge and data engineering, IEEE transactions on**, IEEE, v. 20, n. 5, p. 615–627, 2008. 12

RUDORFF, B. F. T.; ADAMI, M.; AGUIAR, D. A. de; GUSSO, A.; SILVA, W. F. da; FREITAS, R. M. de. Temporal series of evi/modis to identify land converted to sugarcane. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 6., 2009, Cape Town, South Africa. **Proceedings...** Cape Town: IEEE, 2009. p. 252–255. ISBN 978-1-4244-3394-0. 7

RUDORFF, B. F. T.; AGUIAR, D. A.; SILVA, W. F.; SUGAWARA, L. M.; ADAMI, M.; MOREIRA, M. A. Studies on the rapid expansion of sugarcane for ethanol production in São Paulo state (Brazil) using Landsat data. **Remote sensing**, Molecular Diversity Preservation International, v. 2, n. 4, p. 1057–1076, 2010. 7

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Cognitive modeling**, v. 323, p. 214–224, 2002. 31, 38

RUSPINI, E. H.; BONISSONE, P. P.; PEDRYCZ, W. **Handbook of fuzzy computation**. 1. ed. Londres: Institute of Physics Publishing, 1998. 1500 p. ISBN 0-750-30427-8. 26

SAKAMOTO, T.; PHUNG, C. V.; KOTERA, A.; NGUYEN, K. D.; YOKOZAWA, M. Analysis of rapid expansion of inland aquaculture and triple rice-cropping areas in a coastal area of the vietnamese mekong delta using modis time-series imagery. **Landscape and urban planning**, Elsevier, v. 92, n. 1, p. 34–46, 2009. 44, 45

SANDRI, S.; CORREA, C. Lógica nebulosa. **Escola de redes neurais: Conselho Nacional de Redes Neurais**, ITA – São José dos Campos – SP, v. 5, p. 73–90, 1999. 24, 25

SANTANA, O. A.; JÚNIOR, O. A. d. C.; PIMENTEL, C. M. M.; GOMES, R. A. T.; OLIVEIRA, S. N. Modelagem de espectros temporais ndvi-modis, no período de 2000 a 2008, na bacia do rio paracatu, brasil. **Revista brasileira de geofísica**, SciELO Brasil, v. 28, n. 1, p. 47–60, 2010. 8

SHEPPARD, D.; MCPHEE, D.; DARKE, C.; SHRETHRA, B.; MOORE, R.; JUREWITZ, A.; GRAY, A. Predicting cytomegalovirus disease after renal transplantation: suban artificial neural network approach. **International journal of medical informatics**, Elsevier, v. 54, n. 1, p. 55–76, 1999. 31

SILVA, F. B. d.; OLIVEIRA, C. W.; NESS, R. L. L.; ARAÚJO, R. S.; ARAUJO, G. P. Determinação do índice de área foliar (iaf) e do índice de vegetação da diferença normalizada (ndvi) na chapada do araripe através de imagens de satélite (landsat 5 tm), em diferentes anos, na estação seca. **VI encontro nacional da Anppas**, v. 6, p. 168–172, 2012. 6

SKOLE, D.; TUCKER, C. Tropical deforestation and habitat fragmentation in the amazon: satellite data from 1978 to 1988. **Science**, v. 260, p. 1905, 1993. 2

STOEVA, S.; NIKOV, A. A fuzzy backpropagation algorithm. **Fuzzy sets and systems**, Elsevier, v. 112, n. 1, p. 27–39, 2000. 37

TUCKER, C. J. Red and photographic infrared linear combinations for monitoring vegetation. **Remote sensing of environment**, Elsevier, v. 8, n. 2, p. 127–150, 1979. 6

WANG, T.; HUANG, J.-S.; GU, Q.-S. Photometric redshifts of galaxies from sdss and 2mass. **Research in astronomy and astrophysics**, v. 9, n. 4, p. 390, 2009. 10

WARDLOW, B. D.; EGBERT, S. L.; KASTENS, J. H. Analysis of time-series modis 250 m vegetation index data for crop classification in the us central great

plains. **Remote sensing of environment**, Elsevier, v. 108, n. 3, p. 290–310, 2007. 2

WEI, L.; KEOGH, E. Semi-supervised time series classification. In: INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 12., 2006, Philadelphia, PA, USA. **Proceedings...** Philadelphia: ACM, 2006. p. 748–753. ISBN 1-59593-339-5. 11

WOLFE, R. E.; NISHIHAMA, M.; FLEIG, A. J.; KUYPER, J. A.; ROY, D. P.; STOREY, J. C.; PATT, F. S. Achieving sub-pixel geolocation accuracy in support of modis land science. **Remote sensing of environment**, Elsevier, v. 83, n. 1, p. 31–49, 2002. 7

ZADEH, L. A. Fuzzy sets. **Information and control**, Elsevier, v. 8, n. 3, p. 338–353, 1965. 18, 19, 20, 26, 37

ZHANG, J. Developing robust non-linear models through bootstrap aggregated neural networks. **Neurocomputing**, Elsevier, v. 25, n. 1, p. 93–113, 1999. 31

ZHANG, X.; FRIEDL, M. A.; SCHAAF, C. B.; STRAHLER, A. H.; HODGES, J. C. F.; GAO, F.; REED, B. C.; HUETE, A. Monitoring vegetation phenology using modis. **Remote sensing of environment**, Elsevier, v. 84, n. 3, p. 471–475, 2003. 5, 7, 45

ZIMMERMANN, H. J. **Fuzzy set theory**: and its applications. 4. ed. Norwell: Springer, 2001. 174 p. ISBN 8-588-32514-4. 21

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Contam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.