Identificação de Litofácies de Poços de Petróleo Utilizando um Método Baseado em Redes Neurais Artificiais

Elisângela Silva da Cunha

Dissertação submetida à Coordenação do Programa de Pós-Graduação em Informática da Universidade Federal de Campina Grande como parte dos requisitos necessários para obtenção do grau de Mestre em Informática.

Área de Concentração: Ciência da Computação

Herman Martins Gomes - Orientador
Marcelo Alves de Barros - Co-orientador

Campina Grande, Paraíba, Brasil
Elisângela Silva da Cunha, Agosto 2002

CUNHA, Elisângela Silva da

C972I

Identificação de Litofácies de Poços de Petróleo Utilizando um Método Baseado em Redes Neurais Artificiais.

Dissertação (Mestrado), Universidade Federal de Campina Grande, Centro de Ciências e Tecnologia, Coordenação de Pós-Graduação em Informática, Campina Grande - Pb, Agosto de 2002.

108 p. Il.

Orientador: Dr. Herman Martins Gomes

Palavras-Chave:
1. Inteligência Artificial
2. Redes Neurais
3. Perfilagem e Testemunhagem

CDU - ~~007.52~~ 004.8(043)

"IDENTIFICAÇÃO DE LITOFÁCIES DE POÇOS DE PETRÓLEO UTILIZANDO UM MÉTODO BASEADO EM REDES NEURAIS ARTIFICIAIS"

ELISÂNGELA SILVA DA CUNHA

DISSERTAÇÃO APROVADA EM 30.08.2002

PROF. HERMAN MARTINS GOMES, Ph.D
Orientador

PROF. MARCUS COSTA SAMPAIO, Dr.
Examinador

PROF. FRANCISCO VILAR BRASILEIRO, Ph.D
Examinador

PROF. FRANCISCO PINHEIRO LIMA FILHO, Dr.
Examinador

CAMPINA GRANDE – PB

Agradecimentos

Agradeço aos meus pais, Erlei e Eni, por, mesmo à distância, terem me apoiado e incentivado em todos os momentos desta caminhada. Agradeço ao meu amado Fabrício, que sempre esteve ao meu lado, tornando fáceis os momentos difíceis e melhores ainda os momentos de êxito.

Agradeço ao PRH-ANP/MCT que, financiado pelo MCT/CTPETRO através do FNDCT/FINEP, concedeu os recursos materiais e financeiros necessários para o desenvolvimento deste trabalho.

Aos professores participantes da banca examinadora, Marcus Sampaio, Francisco Brasileiro e Francisco Pinheiro, os meus agradecimentos pelas sugestões e críticas que contribuíram para o enriquecimento do trabalho. Ao geólogo João de Deus S. Filho, agradeço pela atenção e por ter se disponibilizado a contribuir com o trabalho.

Agradeço ao meu orientador Herman Martins Gomes, que sempre procurou conduzir o trabalho com paciência e dedicação.

À Wilma de Carvalho, coordenadora do PRH-25, e à Maritza Montoya, pesquisadora visitante do programa, os meus agradecimentos pela insistência nas mudanças no decorrer do trabalho que permitiu aumentar a qualidade do mesmo.

Agradeço à Aninha, Vera e Zeneide, que sempre se mostraram dispostas a auxiliar os alunos da COPIN. À todas as pessoas que de alguma forma contribuíram para a conclusão deste trabalho, o meu muitíssimo obrigada.

Sobretudo, agradeço a Deus por ter me concedido esta graça e permitido que eu concluísse mais uma etapa da minha vida com sucesso.

Resumo

O principal objetivo deste trabalho é propor, implementar e avaliar um método para identificar automaticamente litofácies (unidades litológicas) a partir de dados de perfis e testemunhos de poços de petróleo. A identificação de litofácies é importante para ajudar na determinação da caracterização de um reservatório e na análise da viabilidade econômica de um poço. Um perfil de poço contém informações sobre as rochas sedimentares que ocorrem ao longo de um intervalo de profundidade, usando uma resolução abaixo de um metro, além de informações de porosidade e permeabilidade. A identificação manual de litofácies a partir de perfis de poços, geralmente, consome muito tempo, envolve a análise de grandes volumes de dados e requer conhecimento específico (algumas vezes heurístico). Uma descrição detalhada das unidades litológicas pode ser obtida através de uma análise de testemunho (amostra real da rocha), mas este processo é muito caro e é realizado apenas para alguns poços. Assim, a necessidade de um método computacional para resolver este problema se torna óbvia. O método proposto consiste em utilizar uma abordagem baseada em Redes Neurais para descobrir conhecimento em uma base de dados de perfis e testemunhos. A base de dados foi fornecida pela Agência Nacional do Petróleo (ANP) e contém dados do Campo Escola de Namorado, no Rio de Janeiro. Tentativas anteriores de resolver este problema usando Redes Neurais utilizaram um conjunto muito limitado e genérico de litofácies e usaram dados de apenas 5 poços. Neste trabalho, foram utilizados 8 poços. As principais etapas do método proposto foram implementadas e validadas a partir do conjunto de dados reais. A taxa média de identificação de litofácies ficou em torno de 80 %. Uma solução para o problema só foi possível após a incorporação de uma estratégia para agrupamento prévio das litofácies e tratamento de padrões problemáticos (regiões de conhecimento incerto nos conjuntos de treinamento e de teste).

Abstract

The main objective of this work is to propose, implement and evaluate a method to automatically identify lictofacies (lithological units) from well log and core data of an oil field. This is important since it can help determine whether a well is economically viable or not. A typical well log contains rocks sedimentary information occurring along a wide depth range using a resolution of under a meter, beyond porosity and permeability informations. Manual lictofacies identification from well logs is usually time consuming, involves the analysis of large amounts of data and relies upon very specific (sometimes heuristic) knowledge. A detailed description of the lithological units can be obtained by a core sample analysis, but this is a very expensive process and is made available just to a few wells. Thus, the need of a computational method to solve the above problem becomes obvious. Our method consists of using a neural network approach to perform knowledge acquisition from a database of well logs and core data. The database was provided by the Brazilian Oil Agency (ANP) and contains data from the Namorado oil field in Rio de Janeiro. A previous attempt to solve this problem using neural networks used data from only 5 wells. In this work, we use data from 8 wells. The main modules of the proposed method were implemented and validated from a real data set. The average identification rate was around 80 %. A solution to the problem was only possible after the incorporation of a lictofacies grouping strategies and after dealing with some problematic patterns (regions of uncertain knowledge within the training and test sets).

Conteúdo

Lista de Figuras

Lista de Tabelas

Capítulo 1

Introdução

Este trabalho de dissertação de mestrado foi desenvolvido no contexto do Programa Interdepartamental de Tecnologia em Petróleo e Gás - PRH (25), um programa de formação de profissionais especializados em petróleo e gás que faz parte do Programa de Recursos Humanos da Agência Nacional do Petróleo (ANP) para o setor Petróleo e Gás, (PRH-ANP/MCT).

O PRH-ANP/MCT, voltado para o nível superior, é uma parceria entre o Ministério da Ciência e Tecnologia e universidades brasileiras cujo objetivo é viabilizar e incentivar a criação de programas específicos para formação de pessoal na área de petróleo e gás natural nos níveis de graduação, mestrado e doutorado. Para tanto, o PRH-ANP/MCT concede apoio financeiro (taxa de bancada) às instituições de ensino e bolsas de estudos aos alunos previamente selecionados.

A Universidade Federal da Paraíba, através do PRH (25), recebeu nos últimos 2 anos cerca de 56 bolsas distribuídas pelos mais variados cursos, dentre os quais destacamos o curso de Pós-graduação em Informática, no qual este trabalho foi desenvolvido. De acordo com cada curso existem diferentes especializações. Este projeto se encontra associado à especialização intitulada Engenharia do Conhecimento, cujo principal objetivo é o de formar profissionais especializados no setor de petróleo e gás, buscando suprir deficiências existentes nesse setor, como por exemplo, o desenvolvimento de ferramentas avançadas de informática para inferir informações "escondidas" nos mais diversos bancos de dados.

Diante deste contexto, esta dissertação investiga um método automático para identificação de tipos de rocha de um reservatório de petróleo a partir de dados de perfis elétricos e

testemunhos. Para tal, são utilizadas Redes Neurais Artificiais, máquinas de aprendizagem biologicamente inspiradas.

O presente capítulo procura descrever brevemente as áreas de pesquisa em petróleo e gás contextualizando a dissertação no âmbito destas áreas. Na próxima seção é apresentada uma breve descrição das áreas de pesquisa em petróleo e gás. Por simplicidade, no decorrer desta dissertação, a expressão "petróleo e gás" será substituída por uma expressão mais concisa em que apenas a palavra petróleo aparece. Em seguida, são discutidos os processos de descoberta e exploração de petróleo. Neste capítulo também é introduzida a técnica de Redes Neurais no contexto da Inteligência Artificial e são discutidos os objetivos e relevância do trabalho. Este capítulo é concluído com a apresentação da estrutura da dissertação.

1.1 Áreas de Pesquisa em Petróleo

Existem atividades de pesquisas em diversas áreas tecnológicas ligadas à indústria do petróleo. Alguns exemplos destas áreas são: exploração, reservas e reservatórios, produção, química e avaliação do petróleo, meio ambiente etc. A seguir, é apresentado um breve comentário sobre as áreas citadas acima, não esquecendo que existem outras áreas de pesquisa de igual importância. Estes exemplos foram escolhidos com o principal objetivo de ilustrar a abrangência da pesquisa em relação ao setor petrolífero e situar o presente trabalho neste contexto.

Exploração

A busca por novas tecnologias de exploração visa reduzir o risco exploratório e otimizar a explotação[1] de petróleo. Estas metas podem ser alcançadas através do desenvolvimento de: métodos para a avaliação de sistemas petrolíferos, tecnologias aplicadas à geoquímica de reservatórios, métodos de prospecção direta de hidrocarbonetos, técnicas de quantificação de processos geológicos, técnicas avançadas de sísmica etc. A título de ilustração, apresentamos a seguir dois exemplos de trabalhos desenvolvidos no Brasil na área de exploração: 1) análise do processo de migração de petróleo utilizando um simulador numérico [Filho, 1994]; 2) modelagem e solução para o problema da otimização da locação de poços *offshore* [Souza

[1] Explotar: extrair (proveito econômico) de alguma área, principalmente no tocante aos recursos naturais.

et al., 2001]. Assim como estes, inúmeros trabalhos em busca de um melhor desempenho na exploração de petróleo podem ser encontrados.

Reservas e Reservatórios

A pesquisa tecnológica na área de reservas e reservatórios está relacionada à Geologia e Engenharia de Reservatórios. O principal objetivo é aumentar a recuperação de hidrocarbonetos. Para isso, diferentes áreas de pesquisa são envolvidas, tais como: geofísica de reservatórios, processos de recuperação, simulação numérica de reservatórios dentre outras. O presente trabalho de dissertação se situa na área de reservas e reservatórios, visto que o principal objetivo faz parte da caracterização de reservatórios, porém utilizando uma técnica baseada em Redes Neurais ao invés de abordagens estatísticas.

Produção

As pesquisas na área de produção buscam inovar e otimizar etapas de processos que constituem as atividades de perfuração, completação, estimulação, elevação e produção. Um exemplo de pesquisa nesta área pode ser encontrado em [Bloch et al., 2001] em que são estudados métodos para estimativa da duração da perfuração e completação de poços *offshore* e em [Silva et al., 2001], onde foi feito um estudo das propriedades de um novo fluido de perfuração.

Química e Avaliação de Petróleo

A pesquisa na área de química e avaliação de petróleo, procura realizar análises químicas do petróleo e de seus resíduos. Em [Spinelli et al., 2001] pode ser encontrado um estudo sobre a caracterização das propriedades físico-químicas dos sistemas petróleo/água.

Meio Ambiente

Esta área de pesquisa busca por tecnologias que contribuam para a preservação do meio ambiente. Em [Macêdo, 2001], por exemplo, pode ser encontrado um sistema de prevenção contra incidentes com incêndio e danos ao meio ambiente em área industrial do setor de petróleo. Como um segundo exemplo, no Departamento de Sistemas e Computação da UFPB

está sendo desenvolvido uma biblioteca digital multimídia para gestão de meio ambiente na área de petróleo [Miranda and Baptista, 2001]

1.2 Descoberta e Exploração de Petróleo

Esta seção procura descrever brevemente os principais métodos de prospecção de petróleo e etapas da perfuração de um poço [Thomas, 2001].

1.2.1 Prospecção

A descoberta de petróleo é uma tarefa que envolve um estudo do comportamento das diversas camadas do subsolo através de métodos geológicos e geofísicos que, em conjunto, conseguem indicar o local mais adequado para a perfuração. Em nenhum momento da prospecção pode-se prever onde existe petróleo, o que ocorre é a indicação do local mais favorável à sua existência. Esta fase fornece uma grande quantidade de informações técnicas. A seguir serão apresentados os principais métodos de prospecção de petróleo, a saber, métodos geológicos, métodos geofísicos e métodos sísmicos.

Métodos Geológicos

Com o apoio de aerofotogrametria, fotogeologia e trabalhos de campo, os geólogos elaboram mapas geológicos de superfície e inferem a geologia de subsuperfície a partir destes mapas e de dados de poços. Além disso, os geólogos analisam informações de caráter paleontológico e geoquímico. Desta forma, consegue-se reconstituir as condições de formação e acumulação de hidrocarbonetos em uma determinada região.

Os mapas geológicos de superfície são continuamente construídos e analisados. A aerofotogrametria consiste em fotografar o terreno utilizando-se aviões devidamente equipados, voando com altitude, velocidade e direção constantes. A fotogeologia consiste na determinação de feições geológicas a partir de fotos aéreas, onde dobras, falhas e outras características geológicas são visíveis.

Métodos Geofísicos

Após o esgotamento dos recursos diretos de investigação (métodos geológicos), são adotados métodos indiretos (geofísicos) para prospecção em áreas potencialmente promissoras. A geofísica pode ser vista como o estudo da terra usando medidas de suas propriedades físicas. O objetivo de usar este método é obter informações sobre a estrutura e composição das rochas em subsuperfície.

A gravimetria e a magnetometria, também chamadas métodos potenciais, foram muito importantes no início da prospecção de petróleo por métodos indiretos, permitindo o reconhecimento e mapeamento das grandes estruturas geológicas que não apareciam na superfície. A prospecção gravimétrica para petróleo estuda as variações de densidade em subsuperfície. A prospecção magnética para petróleo tem como objetivo medir pequenas variações na intensidade do campo magnético terrestre.

Métodos Sísmicos

Existem essencialmente dois métodos sísmicos: de refração e de reflexão. O método sísmico de refração foi muito utilizado na área de petróleo na década de 50, mas atualmente sua aplicação é bastante restrita. Atualmente, o método de prospecção mais utilizado na indústria do petróleo é o método sísmico de reflexão. Este método fornece alta definição das feições geológicas em subsuperfície propícias ao acúmulo de hidrocarbonetos.

O levantamento sísmico inicia-se com a geração de ondas elásticas, através de fontes artificiais, que se propagam pelo interior da Terra, onde são refletidas e refratam nas interfaces que separam as rochas de diferentes constituições petrofísicas. Estas ondas retornam à superfície e são captadas por sofisticados equipamentos de registro.

Tanto em terra quanto no mar, a aquisição de dados sísmicos consiste na geração de uma perturbação mecânica em um ponto da superfície e nos registros das reflexões pelos receptores. Na sísmica 2-D os registros das reflexões são realizados por centenas de canais de recepção ao longo de uma linha reta. Geralmente, a distância entre os canais receptores é de 20 metros. Na sísmica 3-D o levantamento dos dados sísmicos é executado em linhas paralelas afastadas entre si de distância igual à distância entre os canais receptores.

O processamento dos dados segue basicamente o mesmo roteiro tanto em 2-D quanto em

3-D. No entanto, na sísmica 3-D o algoritmo de migração possui a flexibilidade de migrar eventos para a terceira dimensão, permitindo que eventos laterais presentes nas seções 2-D sejam migrados para suas respectivas posições verdadeiras em 3-D. Na sísmica 2-D, o conjunto de traços sísmicos é como uma matriz de dados enquanto na sísmica 3-D é um cubo de dados 3-D. A interpretação de dados 3-D é muito mais precisa e facilitada pelo detalhe das informações.

1.2.2 Perfuração de um Poço de Petróleo

A perfuração de um poço de petróleo é a etapa mais cara do processo de exploração e é onde há maior risco econômico. A perfuração é realizada através de uma sonda rotativa composta por diversos equipamentos, cada um responsável por uma determinada função.

As rochas são perfuradas pela ação da rotação e peso aplicados a uma broca existente na extremidade de uma coluna de perfuração. A coluna de perfuração é formada pelos seguintes componentes principais: comandos (tubos de paredes espessas), tubos de perfuração (tubos de paredes finas) e tubos pesados (promovem uma transição de rigidez entre os comandos e os tubos de perfuração). As brocas são equipamentos que têm função de promover a ruptura e desagregação das rochas ou formações.

Os fragmentos da rocha são removidos continuamente através de um fluido de perfuração ou lama. O fluido é injetado por bombas para o interior da coluna de perfuração através da cabeça de injeção, ou *swivel*, e retorna à superfície através do espaço anular formado pelas paredes do poço e pela coluna. Os fluidos de perfuração são misturas complexas de sólidos, líquidos, produtos químicos e, algumas vezes, até gases. Os fluidos são classificados em fluidos à base de água, fluidos à base de óleo e fluidos à base de ar ou de gás.

Ao atingir determinada profundidade, a coluna de perfuração é retirada do poço e uma coluna de revestimento de aço, de diâmetro inferior ao da broca, é descida no poço. O revestimento é realizado para proteger as paredes dos poços. O anular entre os tubos do revestimento e as paredes do poço é cimentado com a finalidade de isolar as rochas atravessadas, permitindo então o avanço da perfuração com segurança. Após a operação de cimentação, a coluna de perfuração é novamente descida no poço, tendo na sua extremidade uma nova broca de diâmetro menor do que a do revestimento para o prosseguimento da perfuração.

O poço é perfurado em diversas fases, cujo número depende das características das rochas

a serem perfuradas e da profundidade final prevista. Geralmente, o número das fases de um poço é de três ou quatro, podendo chegar a oito, em determinados casos. Cada uma das fases é caracterizada pelos diferentes diâmetros das brocas e é concluída com a descida de uma coluna de revestimento e sua cimentação. O objetivo da cimentação é fixar a tubulação e evitar que haja migração de fluidos entre as diversas zonas permeáveis atravessadas pelo poço, por trás do revestimento.

Durante a perfuração de um poço, podem ocorrer várias operações especiais, tais como: controle de *kicks*, operações de pescaria e testemunhagem. *Kick* é a perda de controle do fluido de perfuração. Eventualmente, algum objeto pode cair no poço, ser partido ou ficar preso no poço, impedindo o prosseguimento das operações normais de perfuração. Na indústria do petróleo, este objeto é chamado "peixe", portanto, o termo "pescaria" é aplicado a todas as operações relativas à recuperação ou liberação do "peixe". A operação de testemunhagem é o processo de obtenção de uma amostra real da rocha de subsuperfície e será abordada na Seção 2.4 do Capítulo 2.

Após a perfuração de uma fase do poço, geralmente são descidas várias ferramentas com a finalidade de medir algumas propriedades das rochas, fundamentais para caracterização e avaliação econômica. Este procedimento é denominado perfilagem e será abordado na Seção 2.5 do Capítulo 2.

1.2.3 Avaliação das Formações

A avaliação das formações diz respeito às atividades e estudos que buscam definir a capacidade produtiva e a valoração das reservas de óleo ou gás de uma jazida petrolífera. A avaliação das formações baseia-se principalmente na perfilagem a poço aberto, nos testes de pressão a poço revestido e na perfilagem de produção, além de todas as informações obtidas, antes da perfilagem, referentes ao intervalo de perfuração de interesse. Por se tratar de um dos objetos de estudo deste trabalho, a perfilagem a poço aberto será abordada na Seção 2.5 do Capítulo 2.

Os principais objetivos dos testes de pressão são: identificar os fluidos contidos na formação; verificar a pressão estática (período em que o poço está fechado) e a existência de depleção (queda de pressão do reservatório); determinar a produtividade da formação, dos parâmetros da formação e do dano da formação; e fazer amostragem de fluidos para PVT

(Pressão, Volume e Temperatura). A perfilagem de produção é feita através de perfis corridos após a descida do revestimento de produção e completação inicial do poço, visando determinar a efetividade de uma completação ou as condições de produtividade de um poço.

1.2.4 Completação

Completação é o conjunto de operações destinadas a equipar o poço para produzir óleo ou gás (ou ainda injetar fluidos nos reservatórios), de forma que a operação seja segura e econômica. Por atingir toda a vida produtiva de um poço e envolver altos custos, deve-se fazer um planejamento criterioso das operações de completação e uma análise econômica cuidadosa.

1.3 Redes Neurais e Inteligência Artificial

Não existe um consenso na literatura sobre a definição de Inteligência Artificial. Segundo Russell e Norvig, as definições da IA variam em duas dimensões principais. As quatro primeiras definições dizem respeito ao raciocínio e as restantes estão relacionadas ao comportamento. As definições 1, 2, 5 e 6 medem o sucesso em termos do desempenho humano e as demais dizem respeito ao conceito ideal de inteligência, conhecido como racionalidade [Russell and Norvig, 1995].

1. “O novo esforço emocionante de fazer os computadores pensarem ... máquinas com mentes, no sentido completo e literal”

2. “A automação de atividades que associamos com o pensamento humano, tais como tomadas de decisões, resolução de problemas, aprendizagem ...”

3. “O estudo das faculdades mentais com o uso de modelos computacionais”

4. “O estudo de processos computacionais capazes de perceber, raciocinar e agir”

5. “A arte de criar máquinas que desempenham funções que requerem inteligência quando realizadas por pessoas”

6. "O estudo de como fazer os computadores realizarem tarefas em que, no momento, as pessoas são melhores"

7. "Um campo de estudo que procura explicar e emular comportamentos inteligentes em termos de processos computacionais"

8. "A área da ciência da computação que está preocupada com a automação de comportamentos inteligentes" (Definições obtidas de [Russell and Norvig, 1995]).

Um sistema de IA deve ser capaz de fazer três coisas: 1) armazenar conhecimento, 2) aplicar o conhecimento armazenado para resolver problemas e 3) adquirir novos conhecimentos através da experiência. Uma das principais funções da IA é explorar o conhecimento que deve ser representado de tal forma que seja: capaz de realizar generalizações, compreensível, facilmente modificável e útil em muitas situações.

As técnicas de Inteligência Artificial são usadas para resolver problemas que as técnicas convencionais não conseguem solucionar. A aplicação de técnicas de IA a um determinado problema não garante que será encontrada a melhor solução e sim uma solução aceitável.

A IA pode ser dividida em duas abordagens: IA simbolista e IA conexionista. Estas abordagens podem ser comparadas levando-se em consideração três aspectos principais: nível de explicação, estilo de processamento e estrutura representacional.

Nível de explicação - na IA simbolista, a ênfase está na construção de representações simbólicas. Do ponto de vista cognitivo, a IA supõe a existência de representações mentais e de seus modelos cognitivos como um processamento sequencial de representações simbólicas. Na IA conexionista a ênfase está no desenvolvimento de modelos de processamento distribuído paralelo inspirados no processamento realizado por neurônios dentro do cérebro.

Estilo de processamento - na IA simbolista, o processamento é sequencial. Mesmo quando não existe uma ordem pré-determinada, as operações são realizadas passo-a-passo. Em contraste, na IA conexionista, o paralelismo não só é essencial conceitualmente para o processamento da informação, mas também é a fonte da sua flexibilidade.

Estrutura Representacional - na IA simbolista, a representação simbólica possui uma estrutura quase linguística. Para a IA conexionista a natureza e a estrutura das representações são problemas cruciais.

A IA simbolista pode ser descrita como a manipulação formal de uma linguagem de algoritmos e representações de dados de forma *top-down*, ou seja, parte da generalização em direção aos dados específicos. A IA conexionista pode ser descrita como processadores distribuídos paralelos com uma habilidade natural para aprender e que geralmente operam de forma *bottom-up*, ou seja, dos dados em direção à generalização [Haykin, 1999].

Um dos maiores atrativos dos sistemas conexionistas é que eles empregam representações do conhecimento que parecem ser mais fáceis de serem aprendidos que os sistemas simbólicos. Quase todos os sistemas conexionistas têm um componente de aprendizagem forte. No entanto, geralmente, os algoritmos de aprendizagem de Redes Neurais, principal representante da IA conexionista, envolvem um grande número de exemplos de treinamento e gastam mais tempo no treinamento que as técnicas simbólicas [Rich and Knight, 1991].

Uma das principais desvantagens das Redes Neurais é que o resultado de aprendizagem pode ser de difícil compreensão para um ser humano. Uma vantagem dos modelos conexionistas sobre os sistemas simbólicos é que os conceitos similares possuem representações similares. Nos modelos simbólicos os conceitos similares podem possuir representações diferentes.

Neste trabalho a técnica de Redes Neurais é utilizada devido às suas elevadas capacidades de aprendizagem a partir de exemplos, e de generalização. Maiores detalhes sobre Redes Neurais poderão ser encontrados na Seção 3.1.3 do Capítulo 2. Além disto, é realizada uma revisão bibliográfica de possíveis técnicas para extração de regras de Redes Neurais. Contudo, devido ao tempo disponível para o desenvolvimento desta dissertação, decidimos não incluir a implementação e teste de algoritmos para extração de regras no escopo dos objetivos da dissertação.

1.4 Objetivos e Relevância

O principal objetivo deste trabalho é descobrir padrões em bases de dados relacionadas a exploração de petróleo, mais especificamente, caracterização das unidades litológicas de poços

e reservatórios. O escopo deste trabalho se restringe à identificação das litofácies de poços de petróleo, um problema importante que pode ajudar na caracterização do reservatório e verificação da viabilidade econômica de um poço.

A determinação manual de litofácies é um processo intensivo que envolve o gasto de uma quantidade considerável de tempo por parte de um especialista experiente. O problema se torna muito mais difícil à medida que aumenta o número de perfis simultâneos (que são medidas de determinadas propriedades da formação) a serem analisados.

Uma descrição detalhada da estrutura litológica de um reservatório pode ser obtida através da análise de testemunho (amostra real da rocha). No entanto, este processo é muito caro e é realizado apenas para alguns poços escolhidos estrategicamente. Desta forma, fica evidente a necessidade de um método automático para resolver o problema acima.

O método proposto neste trabalho para resolver o problema de determinação de litofácies é baseado em Redes Neurais Artificiais, visando adquirir conhecimentos a partir de dados de perfis e testemunhos. Como as Redes Neurais não oferecem uma representação do conhecimento facilmente compreensível, temos como atividade complementar a este trabalho a realização de uma revisão bibliográfica sobre extração de conhecimento na forma de regras a partir das redes treinadas.

Para realizar o treinamento da Rede Neural são utilizados dados de perfis e testemunhos de poços de petróleo do Campo Escola de Namorado - Bacia de Campos, Rio de Janeiro, fornecidos pela Agência Nacional do Petróleo (ANP), financiadora deste projeto.

Com respeito à aplicação no setor petrolífero, é de grande importância que sejam desenvolvidos sistemas que forneçam conhecimento útil para auxílio à tomada de decisão.

1.5 Estrutura da Dissertação

Esta dissertação está dividida em sete capítulos. O Capítulo 2 apresenta alguns conceitos fundamentais de Engenharia de Petróleo. O principal objetivo deste capítulo é descrever conceitos considerados necessários para a compreensão do trabalho. As seções abrangem um breve histórico, os constituintes do petróleo, o processo de acúmulo de petróleo e duas técnicas utilizadas nas etapas de perfuração e caracterização de poços de petróleo: testemunhagem e perfilagem.

O Capítulo 3 descreve o processo da descoberta de conhecimento em bases de dados e suas peculiaridades, como representação do conhecimento e técnicas de mineração de dados, dando ênfase à técnica de Redes Neurais Artificiais.

O Capítulo 4 é composto por uma revisão bibliográfica. O objetivo é descrever as principais técnicas utilizadas para identificação de litofácies, extração de regras de Redes Neurais Artificiais, bem como mostrar outras aplicações de Redes Neurais à Indústria do Petróleo. As seções apresentam técnicas estatísticas amplamente utilizadas e técnicas baseadas em Redes Neurais para o reconhecimento de litofácies. Em seguida, são apresentadas algumas abordagens para extração de regras de Redes Neurais Artificiais, como por exemplo: poda da rede treinada, Algoritmos Genéticos e redes baseadas em conhecimento. A seção que descreve as aplicações de Redes Neurais na indústria do petróleo contribuiu para a etapa inicial de definição do problema.

Os Capítulos 5 e 6 são os capítulos centrais da dissertação. O Capítulo 5 apresenta o problema de reconhecimento das litofácies de um reservatório de petróleo. É apresentada a importância do reconhecimento automático de litofácies, bem como um método proposto para esta tarefa. Em seguida, são apresentados os dados utilizados na fase de experimentação do método.

O Capítulo 6 é composto por uma descrição de todos os experimentos desenvolvidos durante o trabalho prático. O capítulo é dividido em seções cuja sequência reflete o processo evolutivo dos experimentos. Também é apresentado o pré-processamento realizado em cada etapa. Por fim, o Capítulo 7 mostra as principais conclusões do trabalho e possíveis trabalhos futuros.

1.6 Sumário

Neste capítulo apresentamos o contexto geral do trabalho e algumas das áreas de pesquisa no setor petrolífero, situando o trabalho na área de reservas e reservatórios. Em seguida, foi descrito o processo sequencial de perfuração de um poço de petróleo. Foi apresentado também o contexto do trabalho do ponto de vista computacional, seguido pelos objetivos e relevância. Por fim, foi apresentada a estrutura da dissertação.

No próximo capítulo, serão abordados alguns conceitos fundamentais da área de petróleo,

importantes para a compreensão do problema a ser resolvido.

Capítulo 2

Introdução à Engenharia de Petróleo

Este trabalho tem caráter interdisciplinar e integra conhecimentos de dois domínios diferentes: Engenharia de Petróleo e a descoberta automática de conhecimento. Assim, para um melhor entendimento do contexto deste trabalho é necessário conhecer alguns conceitos importantes da Indústria do Petróleo como, por exemplo, as etapas envolvidas na perfuração de poços, além do processo de descoberta de conhecimento em bases de dados, que será abordado no próximo capítulo.

A Engenharia de Petróleo envolve o desenvolvimento das acumulações de óleo e gás descobertas durante a fase de exploração de um campo petrolífero, sendo associada, primordialmente, à área de explotação. Os processos que envolvem a transformação do petróleo desde sua descoberta até seu consumo recebe a intervenção de inúmeros especialistas. A Engenharia de Petróleo envolve geólogos, paleontólogos, estratígrafos, sedimentólogos, químicos, geodésicos, geoquímicos, geofísicos, engenheiros mecânicos, elétricos, engenheiros de manutenção, de minas, de perfuração, de completação, de reservatórios, de produção e outros, cada um responsável por uma etapa específica. Apesar da sua multidisciplinaridade, a Engenharia de Petróleo pode ser dividida em quatro áreas básicas: reservatórios, perfuração, completação e produção. Este trabalho envolve dados de perfis e testemunhos, que estão relacionados à caracterização de reservatórios. Nas próximas seções serão apresentados alguns conceitos fundamentais do petróleo e de algumas etapas da fase de perfuração.

2.1 Breve Histórico

O processo de exploração comercial de petróleo começou em 1859, na Pensilvânia, Estados Unidos, logo após a descoberta do Cel. Edwin L. Drake, com um poço de apenas 21 metros de profundidade. Descobriu-se que a destilação do petróleo resultava em produtos que substituíam o querosene obtido do carvão e do óleo de baleia, muito utilizados para iluminação na época [Thomas, 2001].

Com o passar dos anos a prática da busca por petróleo cresceu em todo o mundo. Novas técnicas exploratórias possibilitaram a perfuração de poços com profundidades maiores, bem como o desenvolvimento de estruturas marítimas.

No Brasil, as pesquisas relacionadas ao petróleo começaram em Alagoas em 1891. No entanto, o primeiro poço brasileiro com o objetivo de encontrar petróleo, foi perfurado somente em 1897, por Eugênio Ferreira Camargo, no estado de São Paulo. Este poço atingiu a profundidade final de 488 metros. Alguns anos depois vários poços foram perfurados em diferentes estados, todos sem sucesso [Thomas, 2001].

Em 1941, foi descoberto o primeiro campo comercial, em Candeias, BA. Em 1953, foi criada a Petrobrás, que iniciou a partida decisiva nas pesquisas do petróleo brasileiro [Thomas, 2001]. Atualmente a maior área marítima produtora de petróleo encontra-se no Rio de Janeiro e a terrestre no Rio Grande do Norte.

2.2 O Petróleo

Petróleo é o nome genérico dado às misturas naturais de hidrocarbonetos. Quimicamente, qualquer petróleo é uma mistura extremamente complexa de hidrocarbonetos, outros compostos de carbono e mais algumas impurezas como oxigênio, nitrogênio, enxofre e metais.

Dependendo da temperatura e da pressão a que está submetido, o petróleo pode se apresentar no estado sólido, líquido ou gasoso. Geralmente, o petróleo é encontrado no estado líquido, conhecido como óleo ou óleo cru, ou no estado gasoso, conhecido como gás natural, ou em ambos os estados (parte no estado líquido e parte no estado gasoso), em equilíbrio.

O petróleo no estado líquido é uma substância oleosa, inflamável, menos densa que a água, com cheiro característico e cor variando de acordo com sua origem, oscilando entre o

negro e o âmbar [Popp, 1988]. Ele é constituído por uma mistura de compostos químicos orgânicos. A Tabela 2.1 mostra a análise elementar do óleo.

Hidrogênio	11 - 14%
Carbono	83 - 87%
Enxofre	0,06 - 8%
Nitrogênio	0,11 - 1,7%
Oxigênio	0,1 - 2%
Metais	até 0,3%

Tabela 2.1: Análise elementar do óleo cru típico (% em peso). Obtida a partir de [Thomas, 2001].

A alta porcentagem de carbono e hidrogênio existente no petróleo mostra que os seus principais constituintes são os hidrocarbonetos (compostos químicos orgânicos formados por carbono e hidrogênio).

Independente da origem, todos os petróleos contêm substancialmente os mesmos hidrocarbonetos em diferentes quantidades, o que resulta em diferentes características dos tipos de petróleo. A variação do petróleo de acordo com seus constituintes indica o tipo de derivado produzido: querosene de aviação, diesel, lubrificantes, gasolina, solvente, asfalto etc.

2.3 Geologia do Petróleo

O petróleo tem origem a partir da matéria orgânica depositada junto com os sedimentos. A interação dos fatores - matéria orgânica, sedimentos e condições termoquímicas apropriadas - é fundamental para o início da cadeia de processos que leva à formação do petróleo [Popp, 1988].

Após o processo de geração, é necessário que ocorra a migração e que esta tenha seu caminho interrompido pela existência de algum tipo de armadilha geológica ou trapa, para ter a acumulação do petróleo. A rocha onde o petróleo é gerado é chamada geradora ou fonte e onde se acumula, reservatório. Para que uma rocha se constitua em um reservatório, esta deve apresentar espaços vazios no seu interior (porosidade) que devem estar interconectados, conferindo-lhe a característica de permeabilidade [Popp, 1988].

Uma rocha-reservatório, de uma maneira geral, é composta de grãos ligados uns aos outros por um material, que recebe o nome de cimento. Também existe entre os grãos outro material muito fino chamado matriz. Para que ocorra a acumulação do petróleo, é necessário que alguma barreira se interponha no seu caminho. Esta barreira é produzida pela rocha selante, cuja característica principal é sua baixa permeabilidade. A descoberta de uma jazida de petróleo em uma nova área é uma tarefa que envolve um longo e dispendioso estudo e análise de dados geofísicos e geológicos das bacias sedimentares [Popp, 1988].

2.4 Testemunhagem

A testemunhagem é uma das operações especiais que podem ocorrer durante a perfuração de um poço. A testemunhagem é o processo de obtenção de uma amostra real de rocha de subsuperfície, chamada testemunho, com alterações mínimas nas propriedades naturais da rocha estudada. Com a análise deste testemunho obtém-se informações valiosas sobre a geologia da formação (tais como litologia, textura, porosidade, permeabilidade, saturação de óleo e água etc), que serão utilizadas pela engenharia de reservatórios, completação (conjunto de operações destinadas a equipar o poço para produzir óleo ou gás) e perfuração [Thomas, 2001].

Quando o geólogo quer obter uma amostra da formação que está sendo perfurada, a equipe de sonda coloca uma coroa de testemunhagem no barrilete. A coroa de testemunhagem é uma broca com um furo no meio que permite que a broca corte o testemunho.

O barrilete de testemunhagem é um tubo especial que geralmente mede 9, 18 ou 27 metros. O barrilete, que é onde irá se alojar o testemunho, é colocado na parte interna da coluna de perfuração. Durante a operação, à medida que a coroa avança, o cilindro de rocha não perfurado é encamisado pelo barrilete interno e posteriormente trazido à superfície.

Os testemunhos permitem que os geólogos analisem uma amostra real do reservatório. A partir dessa amostra eles muitas vezes podem saber se o poço será produtivo ou não.

Na testemunhagem com barrilete convencional, ao final de cada corte de um testemunho é necessário trazer a coluna à superfície através de uma manobra, o que aumenta o tempo e o custo da operação. Assim, foi desenvolvida a testemunhagem a cabo, onde o barrilete interno pode ser removido até à superfície sem a necessidade de se retirar a coluna [Thomas,

2001].

Algumas vezes pode haver a necessidade de se testemunhar alguma formação já perfurada. Nestes casos, emprega-se o método de testemunhagem lateral. Neste caso, cilindros ocos, presos por cabos de aço a um canhão, são arremessados contra a parede da formação para retirar amostras da rocha. Ao se retirar o canhão até a superfície, são arrastados os cilindros contendo as amostras retiradas da formação [Thomas, 2001].

2.5 Perfilagem

Após a perfuração de uma fase do poço, geralmente são descidas várias ferramentas com a finalidade de medir algumas propriedades das rochas, fundamentais para caracterização e avaliação econômica. Este processo é conhecido como perfilagem.

A perfilagem permite obter informações importantes a respeito das formações atravessadas pelo poço, tais como litologia (tipo de rocha), espessuras, porosidade, prováveis fluidos existentes nos poros e suas saturações. A maior limitação da perfilagem é a pequena extensão de seu raio de investigação lateral, de modo que apenas a vizinhança do poço é analisada pela perfilagem [Thomas, 2001].

A perfilagem pode revelar a existência de óleo e gás suficientes para justificar os gastos de completação do poço. Esta operação geralmente é feita por serviços terceirizados. Nas sondas terrestres a companhia contratada envia uma unidade de perfilagem montada em um caminhão, enquanto no mar a unidade é fixa na sonda, instalada num pequeno abrigo. A unidade de perfilagem é equipada com computadores, guinchos e controles que executam a operação.

O perfil de um poço é a imagem visual, em relação à profundidade, de uma ou mais características ou propriedades das rochas perfuradas (resistividade elétrica, potencial eletroquímico natural, tempo de trânsito de ondas mecânicas, radioatividade natural ou induzida etc). Tais perfis, obtidos através do deslocamento contínuo de um sensor de perfilagem (sonda) dentro do poço, são denominados genericamente de perfis elétricos, independentemente do processo físico de medição utilizado [Thomas, 2001].

A ferramenta de perfilagem é descida no poço em um cabo condutor até a profundidade desejada. A unidade puxa a ferramenta que sobe pelo poço detectando certos aspectos da

formação por onde ela passa. A informação é enviada à superfície pelo cabo condutor e registrada pelos computadores. O registro é impresso para posterior análise.

Tipos de perfis

Existem vários tipos de perfis utilizados para as mais diversas aplicações, todos com o objetivo de avaliar melhor as formações geológicas quanto à ocorrência de uma jazida comercial de hidrocarbonetos. Os perfis mais comuns são: Potencial Espontâneo, Raios Gama, Neutrônico, Indução, Sônico, Densidade e Caliper.

- **Potencial Espontâneo (SP)**: é o registro da diferença de potencial entre um eletrodo móvel descido dentro do poço e outro fixo na superfície. Este perfil permite determinar as camadas permoporosas, calcular a argilosidade das rochas, determinar a resistividade da água da formação e auxiliar na correlação de informações com poços vizinhos.

- **Raios Gama (GR)**: permite detectar e avaliar a radioatividade total da formação geológica. Utilizado na identificação da litologia, identificação de minerais radioativos e para o cálculo do volume de argilas ou argilosidade. Também pode ser útil para interpretação de ambientes deposicionais e na investigação da subida do contato óleo-água em reservatórios fraturados.

- **Neutrônico (NPHI)**: os perfis mais antigos medem a quantidade de raios gama de captura após excitação artificial através de bombardeio dirigido de nêutrons rápidos. Os mais modernos medem a quantidade de nêutrons epitermais e/ou termais da rocha após o bombardeio. São utilizados para estimativas de porosidade, determinação do volume de argila, pode auxiliar na identificação da litologia e dos fluidos da formação e detecção de hidrocarbonetos leves ou gás.

- **Indução (ILD)**: fornece leitura aproximada da resistividade, através da medição de campos elétricos e magnéticos induzidos nas rochas. A resistividade é a propriedade da rocha permitir ou não a passagem de uma corrente elétrica.

- **Sônico (DT)**: mede a diferença nos tempos de trânsito de uma onda mecânica através das rochas. É utilizado para estimativa de porosidade, identificação de litologia, correlação poço a poço, estimativas do grau de compactação das rochas ou estimativa

das constantes elásticas, detecção de fraturas e apoio à sísmica para a elaboração do sismograma sintético.

- **Densidade (RHOB)**: detecta os raios gama defletidos pelos elétrons orbitais dos elementos componentes das rochas, após terem sido emitidos por uma fonte colimada situada dentro do poço. Além da densidade das camadas, permite o cálculo da porosidade e a identificação das zonas de gás. É utilizado também como apoio à sísmica para o cálculo do sismograma sintético.

- **Caliper**: fornece o diâmetro do poço. É aplicado no cálculo do volume de cimento para tampões ou cimentação do revestimento, apoio a operações de teste de formação, controle de qualidade de perfis e indicações das condições do poço em um determinado intervalo.

Em [Doveton, 1994] podem ser encontrados mais detalhes sobre os tipos de perfis mencionados acima.

2.6 Considerações Finais

Neste capítulo foram discutidos conceitos fundamentais da Engenharia de Petróleo, os quais são importantes para o acompanhamento do restante da dissertação. Foi dada ênfase às etapas de testemunhagem e perfilagem, ocorridas durante e após a perfuração de um poço de petróleo, respectivamente.

Visto que o principal objetivo deste trabalho é descobrir conhecimento a partir de dados de perfis e testemunhos, no próximo capítulo serão apresentados alguns conceitos de descoberta de conhecimento em bases de dados e técnicas de mineração de dados, dando ênfase à técnica de Redes Neurais Artificiais.

Capítulo 3

Descoberta de Conhecimento em Bases de Dados

Devido à rápida evolução tecnológica dos últimos anos, o volume de dados armazenados em meios magnéticos vem aumentando substancialmente. Com tantos dados disponíveis, fica evidente a possibilidade da riqueza de informações contida nos mesmos (eventualmente, mesmo com muitos dados, pode não haver informações implícitas que sejam úteis). No entanto, muitas vezes fica difícil extrair todo o conhecimento implícito nos dados. Para resolver isto é necessário utilizar um sistema que seja capaz de extrair conhecimento implícito em bases de dados e que possa auxiliar na tomada de decisão. Este processo de extração de conhecimento é conhecido como Descoberta de Conhecimento em Bases de Dados (*Knowledge Discovery in Databases* - KDD) e será apresentado no decorrer deste capítulo.

3.1 O Processo de KDD

O termo de Descoberta de Conhecimento em Bases de Dados foi proposto em 1989 para referir-se às etapas do processo computacional que objetiva produzir conhecimentos a partir dos dados e, principalmente, à etapa de mineração de dados em informações [Fayyad et al., 1996]. O processo de KDD é dividido em várias etapas, conforme ilustrado na Figura 3.1.

A primeira fase no processo de KDD é a definição do problema a ser resolvido. Após a definição do problema parte-se para a seleção dos dados apropriados para a análise. Nesta fase ocorre a seleção de atributos relevantes, a discretização de valores de atributos e a re-

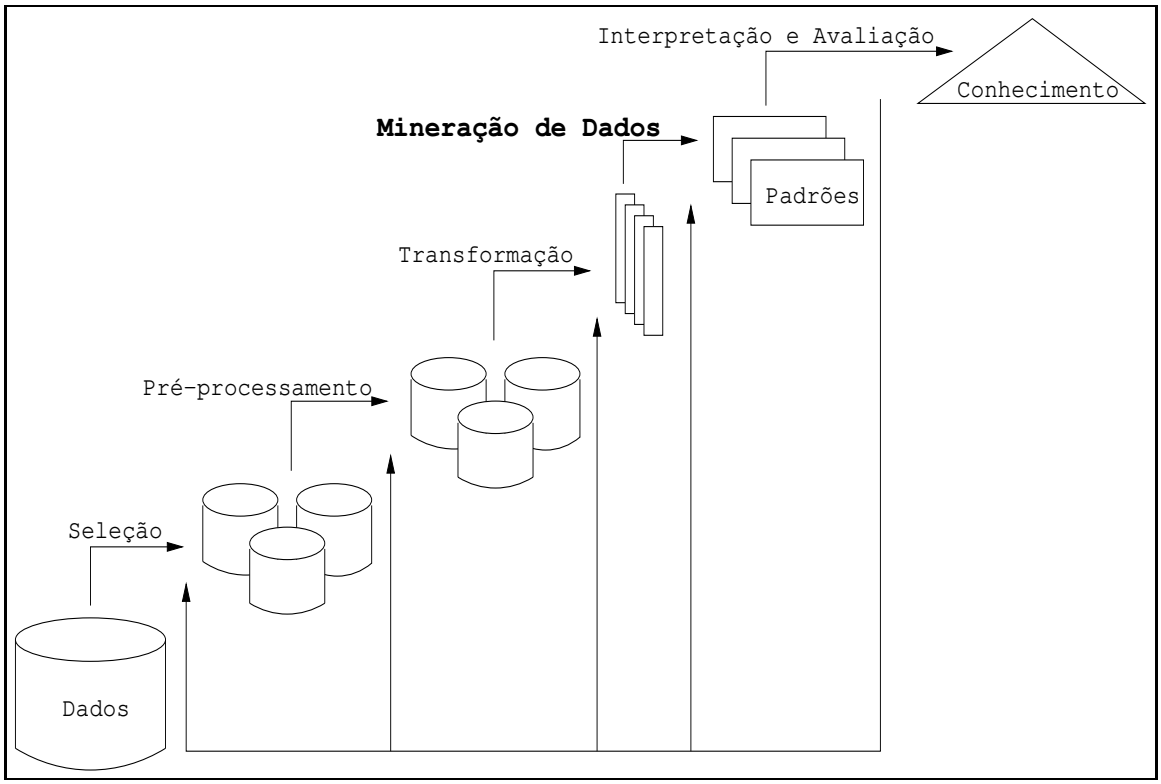

Figura 3.1: Fases do processo de KDD.

moção de inconsistências. A fase seguinte é o pré-processamento dos dados, onde ocorrem vários passos para a construção de uma base de dados consistente, tais como eliminação de ruídos e erros, estabelecimento de procedimentos para verificação da falta de dados e o estabelecimento de convenções para nomeação. A próxima fase é a transformação que tem por objetivo transformar os dados preparados de acordo com o tipo de entrada específica do algoritmo de mineração de dados escolhido para a fase seguinte. Na fase de mineração de dados, é aplicado o algoritmo para descoberta de padrões nos dados. Esta fase envolve a seleção de métodos/técnicas e modelos que melhor se enquadram no cumprimento das metas estabelecidas na identificação do problema [Fayyad et al., 1996]. Ao final desta fase, os padrões gerados devem ser analisados e interpretados para averiguar se o conhecimento gerado é relevante e validá-lo para o seu uso em tomadas de decisões. Esta fase é denominada interpretação. Caso o resultado não seja satisfatório, pode ser necessário repetir parte ou todo o processo de KDD.

A fase de mineração de dados é uma das mais importante do processo de KDD [Fayyad et al., 1996]. Esta fase é a responsável pela extração de conhecimentos implícitos, potencial-

mente úteis e não óbvios dos dados. A tecnologia de mineração de dados possui a vantagem de extrair informação que não seria possível de ser obtida através de consultas tradicionais em bases de dados. As ferramentas tradicionais são limitadas a simples questões feitas pelos usuários, assim, a mineração de dados, extraindo preciosas informações das bases de dados, pode auxiliar o usuário a tomar decisões, melhorando o seu desempenho nos negócios e aumentando a qualidade dos seus serviços.

A tecnologia de mineração de dados pode ser aplicada a diversas áreas, dentre as quais podemos citar: marketing, medicina, economia, engenharia e administração. A mineração de dados pode ser aplicada no ramo de alimentos, na segmentação de mercados, no planejamento da produção industrial, na previsão do volume de vendas, na previsão de mercados financeiros etc [de Carvalho, 2001].

Esta tecnologia também encontra algumas barreiras no seu uso. Algumas delas são: necessidade de grandes volumes de dados armazenados em poderosos servidores[1], complexidade das ferramentas, desafios na preparação dos dados para mineração, dificuldade de realizar uma análise custo/benefício do projeto de descoberta e disponibilidade das ferramentas [OWG, 2000].

No contexto de mineração de dados, conhecer as diferenças entre dado, informação e conhecimento é muito importante. Dado é um conjunto de símbolos que tomado isoladamente não contém nenhum significado claro. Informação é todo dado trabalhado por pessoas ou por recursos computacionais, com valor significativo agregado a ele e com sentido lógico para quem usa a informação. Segundo Nascimento Jr. e Yoneyama [Junior and Yoneyama, 2000], conhecimento é um conjunto de informações que permite articular os conceitos, os juízos e o raciocínio, usualmente disponíveis em um domínio particular de atuação.

Descobrir informações é o mesmo que encontrar padrões nos dados, que geralmente são estruturados, por exemplo, em forma de árvores de decisão e regras SE-ENTÃO. Na subseção seguinte serão descritas algumas formas de representar o conhecimento como padrões estruturados.

[1] Técnicas de amostragem podem ser usadas para minimizar este problema.

3.1.1 Representação do Conhecimento

Existem diversas formas de representar os padrões descobertos pelo algoritmo de mineração de dados. As representações mais utilizadas são: regras de classificação, regras de associação, árvores de decisão e agrupamentos. Para um melhor entendimento destes conceitos será utilizado o exemplo ilustrativo da Tabela 3.1.

Estado	**Temperatura**	**Umidade**	**Vento**	**Jogo**
ensolarado	quente	alta	falso	não
ensolarado	quente	alta	verdadeiro	não
nublado	quente	alta	falso	sim
chuvoso	amena	alta	falso	sim
chuvoso	fria	normal	falso	sim
chuvoso	fria	normal	verdadeiro	não
nublado	fria	normal	verdadeiro	sim
ensolarado	amena	alta	falso	não
ensolarado	fria	normal	falso	sim
chuvoso	amena	normal	falso	sim
ensolarado	amena	normal	verdadeiro	sim
nublado	amena	alta	verdadeiro	sim
nublado	quente	normal	falso	sim
chuvoso	amena	alta	verdadeiro	não

Tabela 3.1: Dados do tempo. Obtida de [Witten and Frank, 1999].

Regras de Classificação

As regras de classificação são construídas em função dos atributos[2] de um conjunto de dados de acordo com o atributo de classificação especificado. O conjunto de regras gerado pelo algoritmo deve ser interpretado em seqüência e pode ser utilizado para classificar novos conjuntos de dados [Witten and Frank, 1999]. Na Tabela 3.1 temos 4 atributos: estado, temperatura, humidade e vento. O atributo de classificação é jogo, onde seus possíveis valores

[2]Cada coluna em uma tabela de dados representa um atributo.

são as classes. Como exemplo de um conjunto de regras de classificação para esses dados temos:

SE estado = ensolarado **E** umidade = alta **ENTÃO** jogo = não

SE estado = chuvoso **E** vento = verdadeiro **ENTÃO** jogo = não

SE estado = nublado **ENTÃO** jogo = sim

SE umidade = normal **ENTÃO** jogo = sim

SENÃO jogo = sim

Existem vários algoritmos que produzem regras de classificação. O método PRISM [Cendrowska, 1987], por exemplo, é uma abordagem que seleciona cada classe do conjunto de treinamento e procura por regras que "cobrem" as instâncias desta classe (*covering algorithm*). Esta abordagem é *bottom-up*, ou seja, começando pelas classes e chegando às regras.

Regras de Associação

As regras de associação definem associações entre diferentes valores de atributos. Geralmente, para um mesmo conjunto de dados, o conjunto de regras de associação é bem maior que o conjunto de regras de classificação. Isto ocorre porque as regras de associação podem prever qualquer atributo e não somente aquele que foi especificado como atributo de classificação. Além disso, as regras de associação podem prever mais de uma situação na mesma regra [Witten and Frank, 1999], conforme o exemplo abaixo.

SE temperatura = fria **ENTÃO** umidade = normal

SE umidade = normal **E** vento = falso **ENTÃO** jogo = sim

SE estado = ensolarado **E** jogo = não **ENTÃO** umidade = alta

SE vento = falso **E** jogo = não **ENTÃO** estado = ensolarado **E** umidade = alta

Vários algoritmos foram desenvolvidos para descobrir regras de associação. O mais utilizado é o Algoritmo Apriori [Agrawal and Srikant, 1994]. Este algoritmo faz uma varredura no arquivo contendo os dados com a finalidade de gerar todos os conjuntos de combinações

de valores de atributos que aparecem no arquivo, com suas frequências. São considerados somente os conjuntos com frequência maior que um determinado valor mínimo, gerando um conjunto L de grandes conjuntos. Após a geração do conjunto L, as regras de associação são extraídas para cada elemento de L.

Árvores de Decisão

As árvores de decisão são representações simples do conhecimento e classificam exemplos em um número finito de classes. Os nós da árvore representam os atributos, os arcos representam os valores para os atributos e as folhas representam as classes. Uma árvore de decisão é geralmente construída de maneira *top-down*, utilizando um algoritmo baseado na aproximação "dividir para conquistar" [Quinlan, 1993]. A Figura 3.2 mostra um exemplo de uma árvore de decisão para os dados da Tabela 3.1.

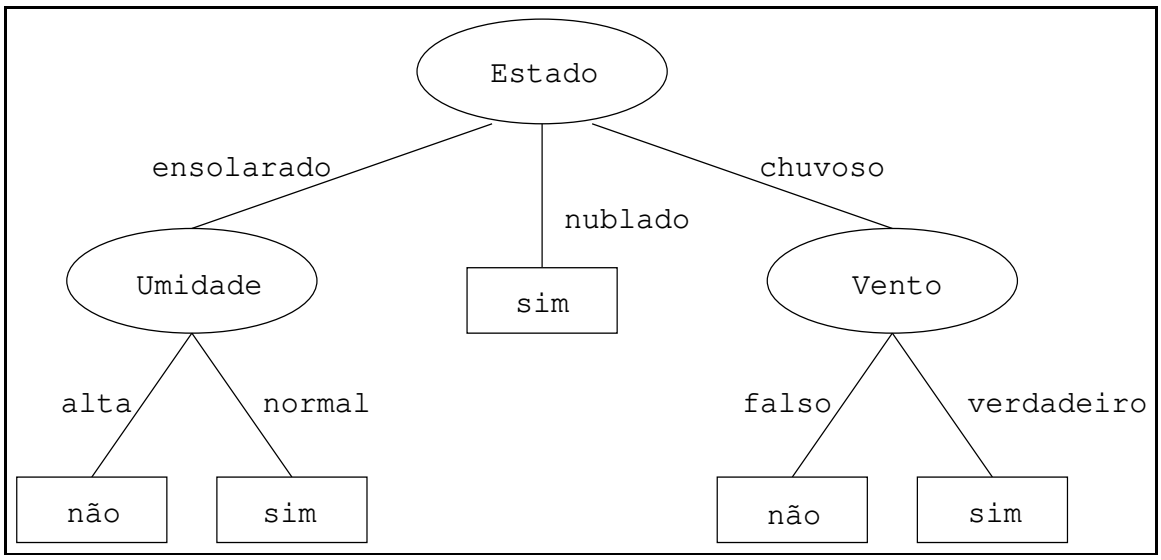

Figura 3.2: Exemplo de árvore de decisão.

O C4.5 [Quinlan, 1993] é um bom exemplo de algoritmo que produz árvores de decisão. Este algoritmo utiliza o ganho de informação para construir a árvore e é descrito com maiores detalhes no Apêndice A por fazer parte do algoritmo de extração de regras investigado neste trabalho.

Agrupamento

O agrupamento/segmentação particiona o banco de dados de forma que os elementos de cada partição ou grupo sejam similares de acordo com algum critério ou métrica e que os elementos de grupos distintos sejam diferentes. Isto pode ser criado estatisticamente ou através da utilização de métodos de indução não-supervisionados, neurais ou simbólicos. O agrupamento/segmentação em banco de dados é o processo de separar o conjunto de dados em componentes que refletem um padrão consistente de comportamento. Neste tipo de representação, um registro pode pertencer a mais de um grupo (interseção de grupos) [Witten and Frank, 1999]. Na Figura 3.3 (a) e (b) é mostrado um exemplo de agrupamento sem e com interseção de grupos, respectivamente.

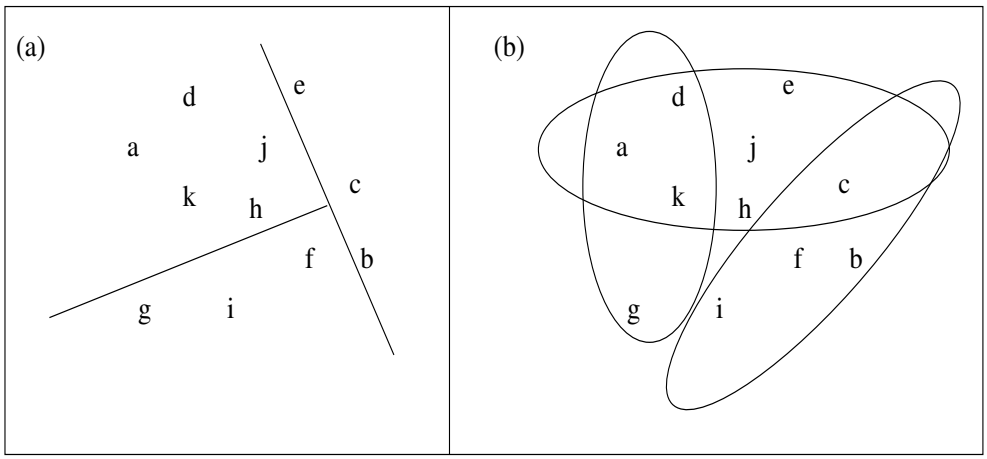

Figura 3.3: Exemplo de agrupamento (a) sem interseção e (b) com interseção.

Em [Alsabti et al., 1998] podem ser encontradas referências para vários algoritmos de agrupamento. O métodos K-means tem mostrado bons resultados na produção de agrupamentos. Neste método, os objetos são colocados em K grupos aleatoriamente. Em seguida, a posição central dos K grupos e o valor da métrica a ser minimizada são determinadas. Um método de otimização global é usado para re-arranjar alguns dos objetos para grupos diferentes. Novos centros e a nova métrica a ser minimizada são determinados. Este procedimento é repetido até que o agrupamento ótimo de objetos seja encontrado [Alsabti et al., 1998; Luke, 2002].

Comparação das Formas de Representação do Conhecimento

Geralmente, as regras de classificação são mais compactas que as árvores de decisão. No entanto, possuem a desvantagem de terem que ser interpretadas em ordem sequencial, o que inviabiliza a possibilidade de modularização. As regras de associação não precisam ser interpretadas em ordem, mas como possuem várias combinações de atributos no seu consequente, o número de regras produzidas é muito maior que o de regras de classificação, o que pode ser um inconveniente. Frequentemente, o agrupamento é seguido por um estágio em que uma árvore de decisão ou um conjunto de regras é inferido para facilitar a interpretação, demonstrando claramente uma desvantagem do agrupamento [Witten and Frank, 1999].

Em termos de clareza da representação, pode haver diferença entre regras e árvores. Em alguns casos, a árvore pode ser muito maior que um conjunto de regras equivalentes. Outra diferença, é que nos casos com múltiplas classes, a divisão da árvore leva em conta todas as classes, tentando maximizar a pureza da divisão, enquanto o método de geração de regras se concentra em uma classe de cada vez, não levando em consideração o que acontece com as outras classes [Witten and Frank, 1999]. Enquanto que as formas de representação do conhecimento discutidas acima são inerentemente simbólicas, a representação do conhecimento em Redes Neurais (técnica de mineração de dados adotada nesta dissertação e discutida em maiores detalhes na próxima seção) é de natureza numérica, em que o conhecimento fica codificado nos pesos das conexões entre neurônios.

Apesar do principal objetivo de Mineração de Dados ser a geração de conhecimento implícito nos dados de forma clara e de fácil interpretação [Holsheimer and Siebes, 1994], em alguns casos este conhecimento não precisa ser representado explicitamente. Por exemplo, para o problema de identificação de litofácies, é muito útil que, mesmo sem conhecer a relação entre os atributos do problema, seja construído um sistema automático para realizar esta tarefa, visto que este tipo de trabalho requer muito tempo de um profissional especializado. Neste caso, o uso de Redes Neurais pode ser mais recomendado que as demais técnicas, pois os dados relativos à este problema geralmente possuem muito ruído e as Redes Neurais têm geralmente apresentado elevada tolerância a ruídos quando comparadas a técnicas simbólicas convencionais [Haykin, 1999].

Existem várias técnicas que podem ser empregadas na fase de mineração de dados, como por exemplo: Indução de Regras, Algoritmos Genéticos e Redes Neurais Artificiais. Na

subseção seguinte será apresentada uma visão geral de tais técnicas.

3.1.2 Técnicas de Mineração de Dados

Existem muitas técnicas que podem ser usadas em mineração de dados e que devem ser escolhidas de acordo com os objetivos do problema a ser resolvido. A Indução de Regras é uma técnica que se refere à detecção de tendências dentro de grupos de dados, ou seja, detecção de regras sobre os dados [OWG, 2000]. Os algoritmos de Indução de Regras produzem regras de classificação ou de associação. Os algoritmos desta técnica fazem parte da IA simbolista.

Algoritmos Genéticos (AG) são modelos estocásticos e probabilísticos de busca e otimização, inspirados na evolução natural e na genética, aplicados a problemas complexos de otimização [Mitchell, 1998]. Em mineração de dados, os AG têm sido empregados para tarefas de classificação e descrição de registros de uma base de dados, além da seleção de atributos de bases de dados que melhor caracterizem o objetivo da tarefa de KDD proposta. Em [Fidelis et al., 2000] podem ser encontradas maiores informações sobre AG em mineração de dados. Na classificação de registros, os modelos de AG geram regras que exprimem uma realidade do domínio de aplicação. Essas regras são de fácil interpretação, o que incentiva o uso dessa técnica [Aurélio et al., 1999]. Os AG podem ser vistos como uma técnica sub-simbólica, pois utilizam uma abordagem baseada na evolução humana mas o conhecimento gerado é representado em forma de regras.

Redes Neurais são uma solução computacional que envolve o desenvolvimento de estruturas matemáticas com a habilidade de aprendizagem. Conforme citado anteriormente, as Redes Neurais possuem representação numérica do conhecimento. Esta técnica faz parte da IA conexionista. Por se tratar do método adotado neste trabalho, a técnica de Redes Neurais será abordada com maiores detalhes na próxima seção. Além das abordagens citadas acima, existem outras que podem ser empregadas para mineração de dados.

3.1.3 Redes Neurais Artificiais

Redes Neurais Artificiais (RNAs) são técnicas computacionais que propõem um modelo matemático baseado na estrutura neural de organismos inteligentes, mais especificamente

o cérebro humano, e tem como principais características: capacidade de aprender, generalização, associação e abstração. Uma RNA procura por relacionamentos, constrói modelos automaticamente e os corrige de modo a diminuir seu próprio erro [Tafner et al., 1995]. A seguir serão apresentados os seguintes aspectos relacionados com Redes Neurais Artificiais: neurônios, arquitetura, aprendizagem, paradigmas, aplicações e extração de regras de RNAs.

Neurônios

O primeiro modelo de um neurônio artificial foi proposto pelo neurofisiologista *Warren McCulloch* e pelo matemático *Walter Pitts* em artigos publicados no ano de 1943 [Tafner et al., 1995].

O cérebro humano é composto por aproximadamente 10 bilhões de neurônios que se comunicam através de sinapses. Sinapse é a região onde dois neurônios entram em contato e através da qual os impulsos nervosos são transmitidos. Juntos, os neurônios formam um grande rede, chamada Rede Neural. A Figura 3.4 mostra um exemplo de um neurônio biológico.

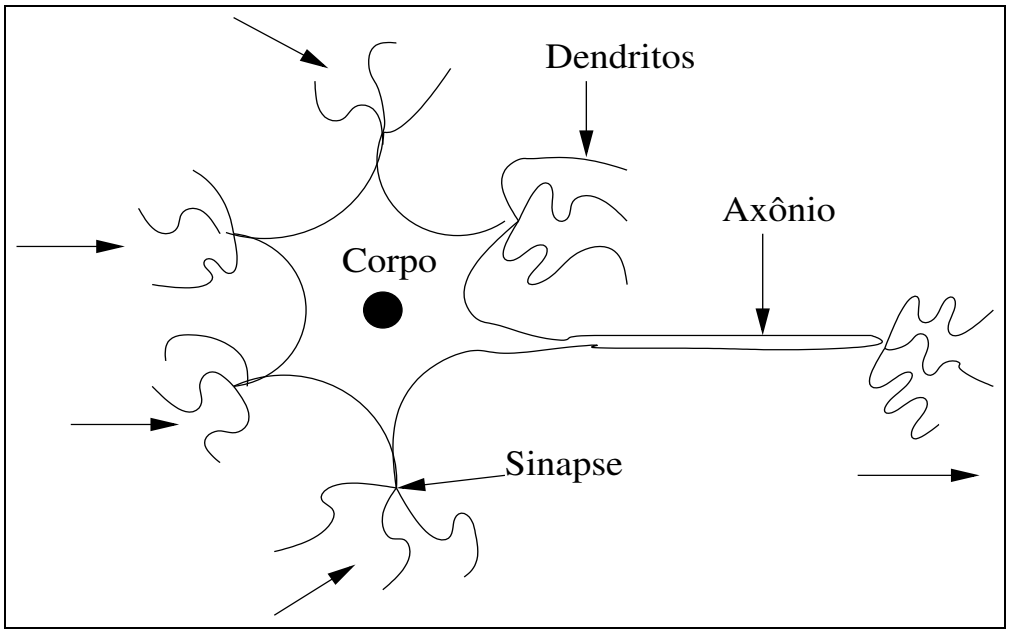

Figura 3.4: Neurônio Biológico.

Os principais componentes do neurônio biológico são: dendritos, responsáveis por receber os estímulos transmitidos pelos outros neurônios; corpo ou soma, responsável por coletar

e combinar informações vindas de outros neurônios; axônio, responsável pela transmissão de estímulos para os outros neurônios.

A transmissão do sinal de uma célula para outra é um complexo processo químico, no qual substâncias específicas são liberadas pelo neurônio transmissor. O efeito é um aumento ou uma queda no potencial elétrico no corpo da célula receptora. Se este potencial alcançar o limite de ativação da célula, um pulso ou uma ação potencial de potência e duração fixa é enviada através do axônio. Neste caso o neurônio está ativo [Tafner et al., 1995; Beale and Jackson, 1990].

O neurônio artificial foi projetado para imitar as características de um neurônio biológico. O neurônio artificial (ver Figura 3.5) possui várias entradas (X_1, X_2, ..., X_p), que podem ser estímulos do sistema ou saídas de outros neurônios. Cada entrada é multiplicada por um peso correspondente (W_1, W_2, ..., W_p), gerando entradas ponderadas. Logo após, todas as entradas ponderadas são somadas e o valor resultante da soma será comparado com um valor limite para ativação do neurônio (função de ativação). Caso o valor da soma alcance o valor limite para ativação do neurônio, ele se ativará, caso contrário, ficará inativo. A saída no neurônio (Y) é o valor processado pela função de ativação.

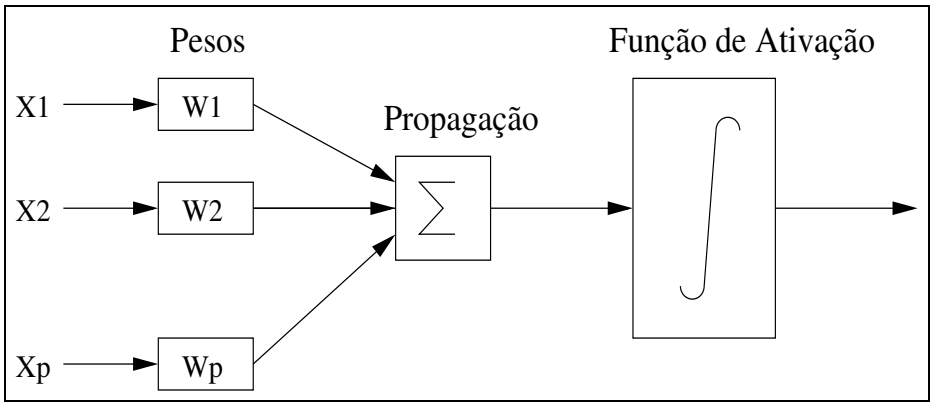

Figura 3.5: Esquema simplificado de um neurônio artificial.

Arquitetura

Combinando diversos neurônios podemos formar o que é chamada de rede de neurônios, ou simplesmente uma rede neural. Os neurônios de uma Rede Neural estão ligados entre si por conexões que comparadas com o sistema biológico, representam o contato do axônio

de um neurônio com o dendrito de outro, formando assim a sinapse. A Figura 3.6 mostra a arquitetura de uma Rede Neural Artificial.

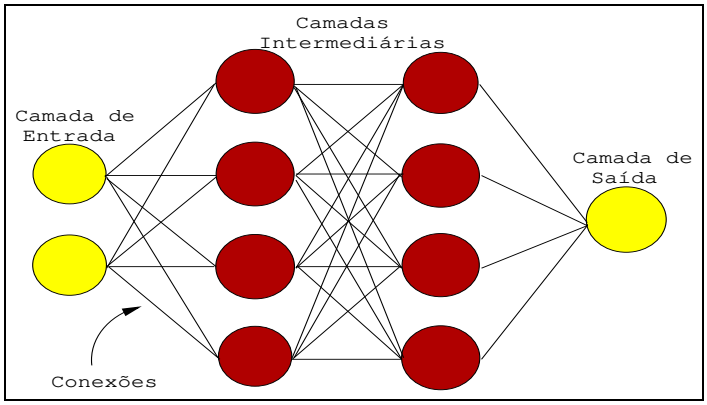

Figura 3.6: Arquitetura de uma Rede Neural Artificial.

As topologias mais comuns de RNAs são as redes recorrentes e as redes não-recorrentes (*feedforward*). As RNAs recorrentes são redes que contêm realimentação das saídas para as entradas. Suas estruturas não são organizadas obrigatoriamente em camadas. As RNAs não-recorrentes são aquelas que não possuem realimentação de suas saídas para suas entradas. Sua estrutura é em camadas, podendo ser formadas por uma ou mais camadas [Beale and Jackson, 1990]. A Figura 3.7 (a) e (b) mostra um exemplo de redes não-recorrentes e recorrentes, respectivamente.

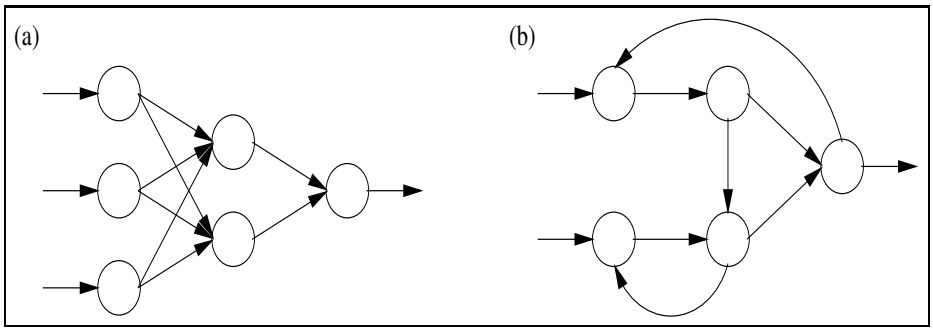

Figura 3.7: Exemplo de uma (a) rede não-recorrente e de uma (b) rede recorrente.

As Redes Neurais com mais de uma camada possuem um conjunto de neurônios de entrada, uma camada de saída e uma ou mais camadas escondidas. A entrada não é considerada uma camada da rede, pois sua única função é receber os padrões de entrada e repassá-los à camada seguinte. Nas camadas intermediárias (normalmente um número de 1 a 2), ocorre a

maior parte do processamento e são consideradas extratoras de características. A camada de saída fornece o resultado da rede.

Aprendizagem

A característica mais importante das Redes Neurais é a capacidade de aprender a partir de exemplos, e de gradualmente melhorar seu desempenho de classificação. Os procedimentos de treinamento que levam as RNAs a aprender determinadas tarefas podem ser classificados em duas categorias: aprendizagem supervisionada e aprendizagem não-supervisionada. Na primeira, um agente externo é utilizado para indicar à rede qual a resposta desejada para o padrão de entrada. Na segunda, não existe agente externo indicando a resposta desejada para os padrões de entrada [Beale and Jackson, 1990].

As Redes Neurais Perceptron são redes de uma única camada, por isso elas são muito limitadas e não servem para representar problemas complexos. Para resolver tais problemas foi criado uma combinação de perceptrons, chamado *Multilayer Perceptron*, que é o modelo utilizado neste trabalho. Para realizar o treinamento da rede será utilizado o algoritmo *Backpropagation* [Beale and Jackson, 1990] mostrado abaixo.

1. Inicializar os pesos e os limiares com valores randômicos pequenos;

2. Apresentar a entrada $X_p = x_0, x_1, x_2, ..., x_{n-1}$ e a saída desejada $T_p = t_0, t_1, t_2, ..., t_{m-1}$ em que n é o número de neurônios de entrada e m é o número de neurônios de saída;

3. Calcular a saída real. Neurônios em cada camada j computam:

$$y_{pj} = f\left[\sum_{i=0}^{n-1} w_i x_i\right] \tag{3.1}$$

   e passam esse valor como entrada para a próxima camada. O valor final da camada de saída é o_{pj};

4. Adaptar os pesos. A adaptação de pesos começa da última camada em direção às camadas anteriores:

$$w_{ij}(t+1) = w_{ij}(t) + \eta \delta_{pj} o_{pj} \tag{3.2}$$

$w_{ij}(t)$ representa os pesos das conexões do neurônio i para o neurônio j no tempo t, η é um termo de ganho e δ_{pj} é um termo de erro para o padrão p no nodo j.

Para as unidades de saída

$$\delta_{pj} = ko_{pj}(1 - o_{pj})(t_{pj} - o_{pj}) \tag{3.3}$$

Para as unidades escondidas

$$\delta_{pj} = ko_{pj}(1 - o_{pj}) \sum_{k} \delta_{pk} w_{jk} \tag{3.4}$$

em que o somatório é feito sobre os k neurônios na camada acima do neurônio j.

Paradigmas

Existem vários modelos de Redes Neurais já publicados em revistas especializadas. Dentre eles podemos destacar como clássicos: *Hopfield, Kohonen, Perceptron, Boltzman e ART* [Beale and Jackson, 1990; Tafner et al., 1995]. Os diversos tipos de Redes Neurais diferem principalmente no tipo de conexão entre os neurônios, no número de camadas e no tipo de treinamento utilizado.

Aplicações

Em mineração de dados, Redes Neurais são aplicadas principalmente para classificação e agrupamento [Aurélio et al., 1999]. A Figura 3.8 mostra um modelo de uma Rede Neural Artificial para mineração de dados. Cada entrada da rede é um registro[3] (tupla), onde cada neurônio recebe o valor de um atributo. A entrada da rede deve ser numérica. Portanto, se existirem atributos nominais, estes deverão ser codificados em uma forma numérica. Cada neurônio de saída corresponde a uma classe. Os círculos do meio representam a camada intermediária (camada escondida). As conexões da rede representam os valores dos pesos.

As taxas de erro de Redes Neurais são equivalentes às taxas de erro de regras produzidas por métodos de aprendizagem simbólicos. As Redes Neurais possuem um desempenho bem melhor quando estão lidando com dados com ruído [Holsheimer and Siebes, 1994].

[3]Cada linha em uma tabela de dados representa um registro.

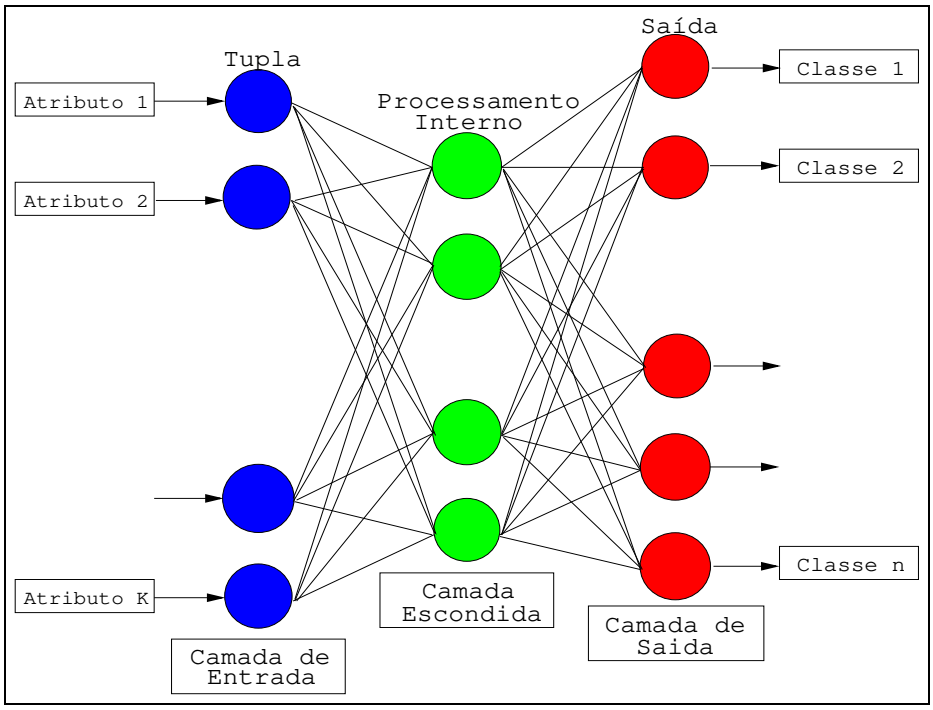

Figura 3.8: Modelo de uma Rede Neural Artificial para Mineração de Dados.

Extração de Regras

A saída de algumas RNAs deve ser transformada em uma forma compreensível humanamente após o processo de treinamento. No entanto, a identificação do conhecimento que relaciona entradas e saídas da rede é um dos problemas encontrados na aplicação de Redes Neurais em mineração de dados. Um outro problema freqüentemente encontrado é o alto tempo de aprendizagem das redes. Devido a estas dificuldades, ainda existem poucas aplicações de Redes Neurais em mineração de dados, pois geralmente os projetistas optam por outras técnicas. Já existem vários estudos sobre como converter os pesos das conexões de uma RNA num conjunto de regras. Algumas abordagens de extração de regras de RNAs serão apresentadas no Capítulo 4.

Redes Neurais não são muito usadas em mineração de dados devido a falta de explicação de seus resultados, ou seja, não há como interpretar as suas saídas relacionando-as com as suas entradas. Devido a isto resolvemos explorar a aplicação de RNAs em mineração de dados mesmo sem resultados compreensíveis humanamente, pois em alguns casos não há necessidade de conhecer a relação entre os atributos do problema. Mesmo assim, sugerimos como trabalho futuro a implementação de um algoritmo para extrair regras de Redes Neu-

rais. Outras técnicas, como por exemplo árvores de decisão, poderiam ser utilizadas para o problema de identificação de litofácies, no entanto os relatos sobre outras aplicações indicam que RNAs tratam melhor dados com ruídos, que é o caso dos dados utilizados neste trabalho.

3.2 Considerações Finais

Neste capítulo foram discutidos conceitos fundamentais do processo de descoberta de conhecimento em bases de dados, os quais são importantes para o acompanhamento do restante da dissertação. Foi dada ênfase à técnica de Redes Neurais.

Um dos principais motivos da escolha da técnica de Redes Neurais para aplicação no problema de identificação de litofácies é que não existem pesquisas nesta área utilizando-se dados de campos brasileiros. Além disto, os trabalhos já realizados utilizaram poucos dados em seus experimentos. Devido a isto, decidiu-se explorar o uso de Redes Neurais para identificação de litofácies utilizando-se dados nacionais e um número maior de poços para treinamento e teste.

Além de técnicas de extração de regras de Redes Neurais, no próximo capítulo serão apresentadas algumas técnicas estatísticas empregadas na identificação de litofácies de um reservatório de petróleo. Também serão apresentadas aplicações específicas de Redes Neurais à indústria do petróleo.

Capítulo 4

Revisão Bibliográfica

O principal objetivo deste capítulo é apresentar uma revisão bibliográfica do estado da arte das técnicas utilizadas na identificação de litofácies, que é o problema central investigado por esta dissertação. Considerando que decidiu-se explorar a técnica de Redes Neurais na resolução deste problema, conforme discutido no Capítulo 2 onde ficou evidente suas características interessantes de aprendizagem a partir de exemplos, alto poder de generalização etc, e que o conhecimento adquirido por uma Rede Neural nem sempre é facilmente interpretado por um ser humano, também incluímos neste capítulo uma revisão das principais técnicas para extração de regras de Redes Neurais Artificiais além de outras aplicações de Redes Neurais à Indústria do Petróleo.

4.1 Técnicas para Identificação de Litofácies

Existem diversas técnicas para determinar as litofácies de poços de petróleo. Nas subseções que se seguem algumas abordagens para esta tarefa serão apresentadas.

4.1.1 Abordagens Estatísticas

Esta seção descreve algumas abordagens estatísticas para determinação de litofácies a partir de testemunhos e perfis de poços. Os dados de perfis e testemunhos carregam informações diferentes sobre a litologia, por isso a determinação de litofácies a partir destas duas fontes de dados é diferente.

A Análise de Agrupamentos (*Cluster Analysis*) é uma das principais abordagens estatísticas para determinar litofácies a partir de testemunhos. O principal objetivo desta abordagem é agrupar objetos similares de acordo com alguma media pré-estabelecida, de forma que objetos não similares pertençam a grupos distintos [Mohn et al., 1987].

Os métodos de projeção são outra técnica de determinação de litofácies a partir de testemunhos. A idéia básica é projetar os dados em dois ou três subespaços dimensionais. Normalmente, estes subespaços são obtidos através de uma Análise dos Componentes Principais (PCA) [Jolliffe, 1986]. A representação gráfica da projeção pode ser inspecionada visualmente.

Ainda se tratando de determinação de litofácies a partir de testemunhos temos o método *box*. Neste caso, cada variável é dividida em um determinado número de intervalos, usando conhecimento geológico e petrofísico. Os intervalos definem um conjunto de caixas p-dimensionais. Uma análise do número de pontos em cada caixa pode revelar estruturas nos dados [Mohn et al., 1987].

Todos estes métodos requerem que as classes de litofácies sugeridas sejam avaliadas por um geólogo. A habilidade destes métodos para formar grupos de profundidade que refletem a litologia do poço depende das variáveis selecionadas. A escolha das variáveis, geralmente, é um resultado de tentativas e erros [Mohn et al., 1987].

Na determinação de litofácies a partir de perfis, serão apresentadas duas abordagens em que os perfis são usados para formar grupos de profundidade que devem ser interpretados comparando-os com os dados de testemunho e uma abordagem em que os perfis são diretamente relacionados às litofácies.

Nos métodos indiretos, os perfis podem servir de entrada para uma análise de agrupamento ou outra técnica mencionada anteriormente. Os grupos obtidos são chamados eletrofácies [Serra, 1989; Lee and Datta-Gupta, 1999]. A interpretação das eletrofácies e a sua correspondência às litofácies deve ser baseada nos dados de testemunho.

A segmentação é outra forma de obter grupos. O principal objetivo desta técnica é dividir os perfis em segmentos homogêneos. A segmentação difere da análise de agrupamento em dois pontos: a) ela leva em consideração a dependência de profundidade dos dados; b) não leva em consideração se segmentos de diferentes partes do poço pertencem ao mesmo grupo [Mohn et al., 1987].

Na abordagem preditiva é feito o cálculo da probabilidade de cada litofácie possível de ocorrer em uma determinada profundidade. As litofácies possíveis devem estar em uma biblioteca e são estabelecidas a partir de testemunhos do poço de interesse ou a partir de outros poços do campo. O método escolhe a litofácie com maior probabilidade e é chamado regra de *Bayes* [Duda et al., 2000].

A última abordagem de determinação de litofácies a partir de perfis é chamada regras de classificação contextual. O principal objetivo desta técnica é utilizar a dependência de profundidades adjacentes, tanto nos perfis quanto na ocorrência de litofácies [Mohn et al., 1987].

4.1.2 Redes Neurais Artificiais Aplicadas à Identificação de Litofácies

Com relação ao uso de RNAs para o reconhecimento de litofácies, foram selecionados dois trabalhos importantes: um de White e colegas [White et al., 1995] e o outro de Siripitayananon e colegas [Siripitayananon et al., 2001].

White e colegas [White et al., 1995] investigaram a viabilidade do uso de RNAs como ferramenta para identificação e reconhecimento de zonas em uma formação heterogênea a partir de perfis de poços. A metodologia adotada consistiu em três passos: 1) coleta de dados, 2) treinamento da RNA para identificação das zonas e 3) verificação. Na fase de coleta de dados foram selecionados poços que tinham dados de perfis de raios gama, indução e densidade, bem como a descrição e análise dos testemunhos. Cinco poços foram identificados como tendo todos os dados necessários, sendo que três foram utilizados para treinamento e dois para teste. Na fase de treinamento, a rede foi alimentada com os dados de perfis (entrada da rede) e com as definições das zonas (saída da rede).

Durante o treinamento foi constatado que a rede não poderia reconhecer as zonas apenas com dados de perfis como entrada. Para resolver este problema, para cada profundidade foi incluído a inclinação da curva com relação ao ponto anterior e posterior. Na fase de verificação, foram apresentados à rede apenas os dados de perfis com suas inclinações para verificar a precisão de predição da mesma. Os resultados mostraram que é possível fazer a identificação de zonas em reservatórios heterogêneos utilizando Redes Neurais, mas poderiam ser melhores se os dados não fossem tão limitados.

Com inspiração neste artigo, nos experimentos realizados nesta dissertação, as incli-

nações das curvas com relação ao ponto anterior e posterior (profundidade), foram utilizadas com o objetivo de atingir bons resultados mais rapidamente. Os resultados dos experimentos descritos em [White et al., 1995] não estavam muito claros, portanto, não foi possível fazer uma comparação entre os resultados do artigo e os obtidos neste trabalho, além da base de dados por eles utilizada não estar disponível publicamente.

Em [Siripitayananon et al., 2001] é proposto integrar dados sísmicos 3-D e dados de poços para fazer a predição de litofácies através de RNAs associadas ao método dos k-vizinhos mais próximos. Foi utilizada uma Rede Neural *feedforward* com aprendizagem supervisionada. A camada de entrada continha 11 nodos, sendo 7 valores de atributos sísmicos, a localização do traço sísmico (coordenadas x e y), *two way time* (TWT - tempo de viagem de uma onda sísmica) para o refletor e um nodo adicional para aumentar o grau de liberdade da rede. A camada escondida era composta por 7 nodos, totalmente conectados à entrada e à saída. A camada de saída tinha 4 nodos representando 4 categorias de litofácies denotadas por um código de 4-bits.

A abordagem proposta consistiu de cinco passos: 1) preparação de dados para treinamento, 2) treinamento da rede usando os dados preparados, 3) predição de litofácies, 4) converção TWT para profundidade e 5) impressão das litofácies para a área de estudo. Foi empregado um algoritmo de classificação dos k-vizinhos mais próximos, visando obter a melhor combinação entre os dados sísmicos e as litofácies correspondentes, obtendo, assim, o conjunto de treinamento.

Foi utilizada a técnica de validação cruzada, onde de cinco poços, foi excluído um de cada vez para teste e os restantes foram usados para o treinamento. A melhor taxa de reconhecimento registrada foi de 81,46 %.

Neste trabalho de dissertação optou-se pelo uso de dados de perfis ao invés de dados sísmicos devido à dificuldade de interpretação e tratamento de dados sísmicos. Siripitayananon e colegas mostraram que é possível estabelecer um relacionamento entre atributos sísmicos e litofácies, embora esta relação seja complicada e aparentemente impossível.

4.1.3 Análise Crítica

As abordagens estatísticas são utilizadas há muito tempo para fazer a predição de litofácies. No entanto, elas requerem que seus resultados sejam avaliados por um especialista. Desta

forma, a disponibilidade de um profissional especializado se torna indispensável. Um outro problema das abordagens estatísticas é o número de parâmetros selecionados, que geralmente resultam de um processo de tentativas e erros.

Devido a estes problemas resolvemos investigar o uso de Redes Neurais para identificação de litofácies ao invés de utilizar uma das abordagens geralmente adotadas. Este trabalho se inspirou em duas pesquisas realizadas nesta área, conforme citado na Subseção 4.1.2. A principal diferença entre este trabalho e os já existentes está na quantidade de dados, bem como na origem dos mesmos. Nos experimentos deste trabalho, foram utilizados dados nacionais de 8 poços. Com o uso de Redes Neurais para a identificação de litofácies, a interpretação de um profissional especializado não se faz necessária e os parâmetros utilizados são sempre os mesmos.

4.2 Técnicas para Extração de Regras de Redes Neurais Artificiais

Conforme citado na Seção 3.1.3, um dos principais problemas do uso de Redes Neurais em mineração de dados é a dificuldade em representar explicitamente na forma de regras o conhecimento gerado pelas mesmas, dificultando assim sua interpretação. A seguir serão descritas algumas abordagens que procuram solucionar este problema.

4.2.1 Rede Neural Baseada em Conhecimento

Towell e Shavlik [Towell and Shavlik, 1993] propuseram um algoritmo para extração de regras refinadas de Redes Neurais baseadas em conhecimento. O método proposto extrai eficientemente regras simbólicas de Redes Neurais treinadas. O algoritmo é dividido, basicamente, em três passos.

O primeiro passo é inserir conhecimento, que não precisa ser nem correto nem completo, em uma Rede Neural especial chamada de KBANN (*Knowledge-Based Artificial Neural Network*). Este passo muda a representação de regras simbólicas para conexionista, fazendo com que as regras se tornem refináveis por métodos de aprendizagem neurais padrão.

O KBANN é uma representação de conhecimento simbólico em Redes Neurais, definin-

do a topologia e os pesos das conexões da rede. Considere como exemplo a base de conhecimento da Figura 4.1 (a), que define os membros da categoria A. A Figura 4.1 (b) representa a estrutura hierárquica destas regras: linhas sólidas e pontilhadas representam as dependências necessárias ou proibitórias, respectivamente. A Figura 4.1 (c) representa a Rede Neural baseada em conhecimento que resulta da tradução desta base de conhecimento em uma Rede Neural. As unidades X e Y na Figura 4.1 (c) são introduzidas na Rede Neural para representar a disjunção no conjunto de regras. Cada uma das outras unidades representam um antecedente ou um conseqüente na base de conhecimento. As linhas grossas representam as ligações com pesos altos, que correspondem às dependências. As linhas finas representam as ligações adicionadas à rede para permitir um refinamento adicional da base de conhecimento.

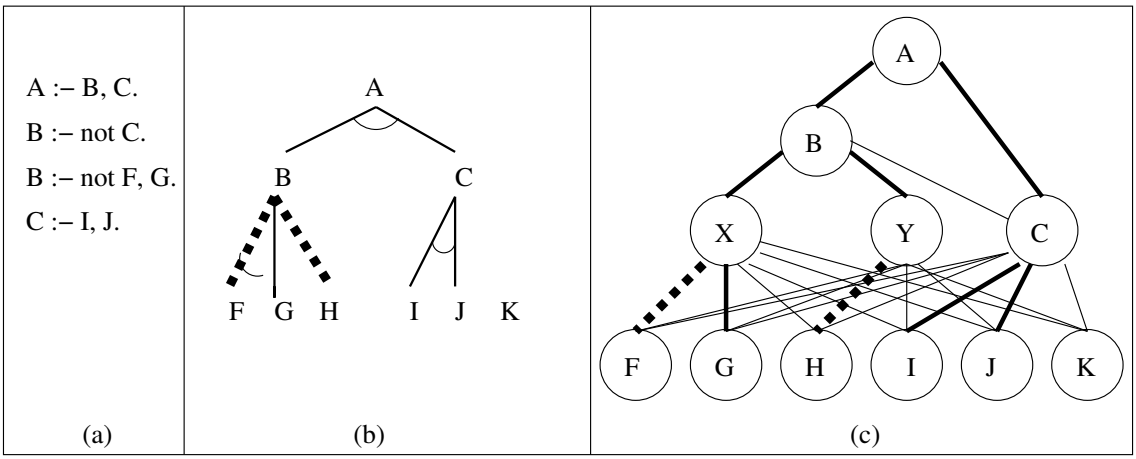

Figura 4.1: Tradução de uma base de conhecimento em uma Rede Neural baseada em conhecimento. Adaptada de [Towell and Shavlik, 1993].

Este procedimento para inicializar as Redes Neurais tem duas vantagens principais: o algoritmo indica as características de entrada que ele julga importantes para a classificação de exemplos e especifica as características derivadas importantes (B e C na Figura 4.1 (b)), auxiliando assim a escolha do número de unidades escondidas na rede.

O segundo passo do algoritmo é treinar a rede usando um conjunto de exemplos de treinamento classificados e um algoritmo de aprendizagem neural padrão, *backpropagation* ou qualquer outro método para otimização de pesos de Redes Neurais *feedforward*. O terceiro e último passo é extrair as regras da rede treinada.

O algoritmo de extração de regras proposto por Towell e Shavlik chama-se MofN. Este algoritmo produz regras do tipo: *Se (M dos N antecedentes são verdadeiros) então (decisão*

de classificação)

O algoritmo é composto de seis passos: agrupar as ligações com pesos similares; ajustar os pesos de todas as ligações, de cada grupo, para a média dos pesos do grupo; eliminar qualquer grupo que não seja significante para a classificação correta dos exemplos; otimizar a rede; formar as regras; e simplificar as regras.

Os principais resultados dos testes deste método mostraram que as regras extraídas reproduzem aproximadamente a precisão da rede da qual elas são extraídas; são superiores às regras produzidas por métodos que refinam diretamente regras simbólicas; são superiores àquelas produzidas por técnicas mais antigas para extração de regras de Redes Neurais treinadas; são "humanamente compreensíveis". Este é um método que procura integrar abordagens simbólicas e conexionistas.

Este método também tem algumas limitações: o método MofN requer que a rede seja baseada em conhecimento, ou seja, que as redes sejam compreensíveis inicialmente; grandes mudanças no significado das unidades, como resultado do treinamento, podem produzir regras de difícil compreensão.

4.2.2 Poda da Rede Treinada

Em [Lu et al., 1995; Lu et al., 1996] é proposto um algoritmo, denominado RX, para extrair regras de Redes Neurais que consiste, basicamente, em três passos:

1. Treinamento da rede: esta fase visa obter o melhor conjunto de pesos para a rede, permitindo que a rede classifique as instâncias de entrada com um nível satisfatório de precisão.

2. Poda da rede: visa remover as conexões e neurônios redundantes sem aumentar a taxa de erro de classificação.

3. Extração de regras: extrai regras de classificação da rede podada.

No treinamento da rede, o número de neurônios de entrada corresponde a dimensionalidade das instâncias de entrada e o número de neurônios de saída é igual ao número de classes. A abordagem adotada para determinar o número de neurônios da camada escondida é começar com muitos neurônios e depois podar as conexões e neurônios redundantes. Foram

utilizadas duas funções de ativação: função tangente hiperbólica, na camada escondida, e função sigmóide, na camada de saída. Visando minimizar os pesos das conexões irrelevantes, facilitando assim a poda, foi utilizada uma função de entropia cruzada (*cross-entropy*) como função de erro. Em [Lu et al., 1995; Lu et al., 1996] o algoritmo de aprendizagem foi uma variação do método quasi-Newton para minimização de funções. Em [Setiono and Liu, 1995] o mesmo algoritmo de extração de regras é utilizado com um treinamento do tipo *backpropagation*.

O algoritmo de poda que visa remover algumas conexões sem afetar a precisão de classificação da rede pode ser encontrado em detalhes em [Lu et al., 1995].

O algoritmo de extração de regras é composto pelos seguintes passos:

1. Discretização dos valores de ativação dos neurônios da camada escondida via agrupamento. Isto possibilita estabelecer uma dependência entre os valores dos neurônios da camada escondida e os valores da camada de saída e uma dependência entre os valores de ativação dos neurônios da camada escondida e os valores da camada de entrada. A partir desta dependência, regras podem ser geradas.

2. Enumerar os valores de ativação discretizados. Gerar regras que relacionam os valores de ativação das unidades escondidas aos valores de saída.

3. Enumerar os valores de entrada que influenciam os valores discretizados dos neurônios da camada escondida que aparecem nas regras do passo anterior. Gerar regras como no passo anterior.

4. Gerar regras que relacionam os valores de entrada e os valores de saída através da substituição das regras, baseado nos 2 passos anteriores.

As regras geradas são do tipo "Se $(a_1\ \theta\ v_1)$ e $(a_2\ \theta\ v_2)$ e ... e $(a_n\ \theta\ v_n)$ então C_j", em que a_i's são os atributos, v_i's são constantes, θ's são operadores relacionais (=, <=, >=, <>) e C_j é uma classe.

Em [Setiono, 2000] é proposto o mesmo algoritmo, com denominação MofN3, com o acréscimo de um passo extra ao algoritmo de extração de regras. O último passo do algoritmo MofN3 é substituir as condições das regras geradas por condições M-de-N (*M-of-N*). Neste

caso as regras finais são do tipo: *Se (M dos N antecedentes são verdadeiros) então* C_j, em que C_j é uma classe.

Em [Setiono and Leow, 2000; Setiono and Leow, 1999] é proposto um método para extração de regras simbólicas de Redes Neurais *feedforward* treinadas com uma única camada escondida, chamado FERNN. O método não requer poda da rede, portanto não é necessário retreiná-la. Assim, a velocidade do processo de extração de regras aumenta consideravelmente, tornando as Redes Neurais uma ferramenta atrativa para geração de regras de classificação simbólicas.

O FERNN consiste em dois componentes principais: um algoritmo de treinamento da rede que minimiza uma função de erro de entropia cruzada (*cross-entropy*) aumentada por uma função de penalização (isto garante que as conexões irrelevantes tenham pesos muito pequenos, portanto estas conexões podem ser removidas sem afetar a precisão de classificação da rede) e um algoritmo de geração de árvores de decisão que gera uma árvore classificadora usando os valores de ativação dos neurônios escondidos da rede. O algoritmo FERNN se encontra no Apêndice A.

Após a árvore de decisão ter sido gerada, os neurônios cujos valores de ativação fazem parte da árvore são considerados relevantes e os demais irrelevantes. Foi desenvolvido um critério simples para remover as conexões consideradas irrelevantes conectando os neurônios de entrada aos da camada escondida, sem afetar a precisão de classificação da rede. Se as conexões satisfazem este critério elas são eliminadas. Quanto mais conexões são removidas, mais simples serão as regras geradas.

Depois de remover as conexões redundantes, as condições da árvore de decisão são reescritas em termos das entradas da rede que não foram removidas. As condições podem ser reescritas como regras DNF (*Disjunctive Normal Form*) ou regras MofN. As regras DNF são expressas como uma disjunção de conjunções, enquanto as regras MofN são expressas como:

> Se {pelo menos/exatamente/no máximo} M das N condições (C1,C2,...,CN)
> são satisfeitas, então ...

O algoritmo foi testado em 15 problemas diferentes e mostraram que o tamanho e a precisão da árvore gerada são comparáveis às regras extraídas por outros métodos que podam

e retreinam a rede.

4.2.3 Algoritmos Genéticos para Extração de Regras

Algoritmos Genéticos (AG) são métodos estocásticos de otimização de problemas que derivam seu comportamento de uma metáfora do processo evolutivo natural. As possíveis soluções de um problema são combinadas e alteradas através de mecanismos inspirados na seleção natural, permutação (*crossover*) e mutação genética. Uma das principais características dos AG é que eles obtém um conjunto de soluções ao invés de uma única solução. Cada possível solução de problema é representada por um cromossomo, que funciona de modo a emular e modelar cromossomos biológicos. Um conjunto de soluções é chamado de população de cromossomos. A população inicial é gerada aleatoriamente. Após a geração da população inicial são aplicadas algumas funções aos cromossomos. A função de "*fitness*" é uma medida de aptidão dos cromossomos para solucionar o problema em questão. A população de cromossomos é modificada a cada geração através de operadores genéticos, permutação e mutação. Desta forma, surgirá uma nova população que passará pelo mesmo processo. Este ciclo ocorrerá até chegar a um resultado satisfatório [Mitchell, 1998].

Alguns pesquisadores adotaram a utilização de Algoritmos Genéticos (AG) para a extração de regras em Redes Neurais. [Nievola et al., 1999] propuseram uma adaptação do algoritmo RX (algoritmo proposto por [Lu et al., 1995]) em conjunto com Algoritmos Genéticos para a extração de regras de Redes Neurais. Os Algoritmos Genéticos foram utilizados para encontrar uma topologia apropriada para uma Rede Neural que permitisse a extração de um conjunto de regras com maior taxa de acerto e menor complexidade.

As regras extraídas são do tipo SE-ENTÃO, em que a parte SE especifica um conjunto de condições sobre valores de atributos previsores e a parte ENTÃO especifica um valor previsto para o atributo classe.

O primeiro passo do algoritmo de extração de regras é discretizar os valores de ativação dos neurônios da camada escondida via agrupamento. O segundo passo é enumerar os valores de ativação de cada neurônio na camada escondida para computar a saída da rede e gerar regras onde o antecedente contém valores da camada escondida e o conseqüente contém valores da camada de saída. O terceiro passo é usar um processo similar ao segundo passo para extrair regras onde o antecedente contém os valores da camada de entrada e o

conseqüente contém os valores da camada escondida. O último passo é combinar os dois conjuntos de regras, gerados nos passos dois e três, criando um conjunto de regras onde o antecedente contém os valores da camada de entrada e o conseqüente contém os valores da camada de saída [Santos et al., 2000].

A diferença entre esta abordagem e o algoritmo RX é que na aplicação de Algoritmos Genéticos não é feita a poda da rede. Ao invés de avaliar regras individuais, os Algoritmos Genéticos avaliam a qualidade do conjunto de regras como um todo, o que é visto como uma vantagem desta abordagem. As regras são avaliadas através de um conjunto de validação que é independente do conjunto de treinamento e do conjunto de testes. O conjunto final de regras é avaliado com base nos dados de teste.

Primeiramente, para testar esta abordagem, foram utilizadas duas bases de dados disponíveis publicamente, cujas taxas de acerto do conjunto de regras no conjunto de exemplos de teste foram de 89,28% e 78,94%. As bases de dados utilizadas foram: *Hayes-Roth* - base com dados de pessoas como, por exemplo, idade e estado civil, com o objetivo de classificar as pessoas em categorias; e *Zoo* - base com dados de animais, com o objetivo de classificar o tipo do animal. A compreensibilidade do conjunto de regras foi de 0,92 e 0,946 numa escala de 0 a 1. A medida de compreensibilidade (CP) é calculada em função do número de regras e do número de condições das regras, conforme a Equação 4.1. A compreensibilidade aumenta à medida que o número de regras e condições diminui [Nievola et al., 1999].

$$CP = 1 - \frac{2 * \frac{R}{MAX\_R} + \frac{C}{MAX\_C}}{3} \tag{4.1}$$

em que R é o número de regras, C é o número de condições, MAX_R é o maior número de regras extraídas até o momento e MAX_C é maior número de condições extraídas até o momento.

Posteriormente, foram utilizadas mais três bases de dados disponíveis publicamente. As bases utilizadas foram: *Iris* - base com dados de plantas com o objetivo de classificar os tipo da planta íris; *Wine* - base com dados de análises químicas de vinhos com o objetivo de determinar a origem das bebidas; *Monks1* - base com dados de pessoas para determinar se são monges ou não. Para estas bases de dados foi utilizada a metodologia de validação cruzada, com um fator de validação igual a 5, ou seja, o conjunto de dados foi particionado

em 5 subconjuntos e o algoritmo de extração de regras foi aplicado 5 vezes, cada uma destas vezes um subconjunto de dados foi usado como teste e os 4 restantes, como treinamento. Após o processo de validação cruzada ter sido realizado, as 5 partições foram unidas em um único conjunto de dados de forma a gerar as regras finais a serem apresentadas ao usuário [Santos et al., 2000].

Um outro trabalho foi desenvolvido por pesquisadores da Universidade Federal do Rio de Janeiro [Hruschka and Ebecken, 2000], em que onde foi proposto um algoritmo genético de agrupamento para extração de regras de Redes Neurais Artificiais treinadas. A metodologia utilizada foi baseada no agrupamento de valores de ativação das unidades escondidas. O algoritmo de extração de regras consiste basicamente em 2 passos:

1. Aplicar um algoritmo de agrupamento para encontrar grupos de valores de ativação das unidades escondidas para cada classe;

2. Enumerar, para cada unidade escondida, os valores de entrada predominantes e gerar um conjunto de regras para descrever os valores discretizados das unidades escondidas em relação às entradas.

O algoritmo desenvolvido foi avaliado experimentalmente na *Australian Credit Approval Database*, uma base de dados disponível publicamente que trata de aplicações de cartões de crédito. Várias Redes Neurais *backpropagation* foram treinadas e a que apresentou a menor taxa de erro foi escolhida para extração de regras. O conjunto de regras obtido produziu uma taxa de classificação média igual a 72,61 %. Por um lado, os resultados mostraram que o método é bastante promissor e, por outro, mostraram que é possível obter resultados melhores aplicando diferentes heurísticas para ajustar a população inicial do Algoritmo Genético. Além disso, é possível encontrar uma função de *fitness*[1] melhor, assim como é necessário encontrar melhores parâmetros para o algoritmo genético de agrupamento [Hruschka and Ebecken, 2000].

Antes de propor este algoritmo, [Hruschka and Ebecken, 1999] já haviam proposto um algoritmo de extração de regras baseado no RX [Lu et al., 1995]. Ao ser comparado com

[1] *Fitness* é a heurística que determina em que medida uma solução é boa ou não, ou seja, o grau de adaptação de um indivíduo ao problema [Pinto and Monteiro, 1998].

uma árvore de classificação este apresentou uma medida de complexidade inferior, mas uma taxa de classificação relativamente menor.

4.3 Outras Aplicações de Redes Neurais na Indústria de Petróleo

Redes Neurais Artificiais (RNAs) podem auxiliar engenheiros de petróleo na resolução de alguns problemas fundamentais da indústria, como por exemplo predição da permeabilidade da formação, bem como problemas específicos que não podem ser resolvidos por métodos convencionais [Mohaghegh and Ameri, 1995].

RNAs têm sido usadas para reconhecimento de litologias e minerais a partir de perfis de poços, estimação de reserva mineral e processamento de dados sísmicos [Huang et al., 1996]. A permeabilidade é uma das características mais importantes da formação dos hidrocarbonetos. A estimativa de permeabilidade é necessária para predição dos volumes e caminhos da migração dos fluidos, bem como para o sucesso do projeto e gerenciamento de processos no desenvolvimento de campos de óleo e gás. Frequentemente, a permeabilidade é medida a partir de testemunhos ou avaliada a partir de testes de formação, ambos disponíveis apenas para alguns poços devido ao custo excessivo.

É muito comum traçar gráficos de porosidade versus permeabilidade para vários poços e gerar a correlação entre estas variáveis para estimar a permeabilidade da formação em outros poços que não foram testemunhados, mas isto é válido apenas para formações homogêneas, que são raras na maioria das aplicações práticas. Quando o reservatório é heterogêneo esta técnica perde a credibilidade [Mohaghegh et al., 1995]. Apesar de complexo e difícil de expressar existe um relacionamento entre a permeabilidade e os perfis dos poços [Huang et al., 1996].

Huang e colaboradores [Huang et al., 1996] criaram uma Rede Neural para predizer/estimar permeabilidade a partir de dados de perfis, profundidade que os dados foram medidos e coordenadas geográficas dos poços. Eles utilizaram uma variação do algoritmo *Backpropagation*, o *Quick-Prop*, que é mais rápido no processo de treinamento e mais robusto no estágio de aprendizagem.

No projeto apresentado foram usados dados de 4 poços para treinamento e de 1 poço

para teste. As variáveis de entrada da rede foram as coordenadas geográficas (latitude e longitude), profundidade dos dados medidos, potencial espontâneo, raios gama, densidade, sônico, porosidade neutrônica e correção da densidade. Foram selecionados 213 casos de treinamento, 73 casos para o conjunto de supervisionamento e 253 casos para o conjunto de teste. Todas as entradas foram linearmente normalizadas para o intervalo [0, 1]. Uma transformação logarítmica foi aplicada aos valores de permeabilidade e estes resultados foram linearmente normalizados para o intervalo [-0.5, 0.5] que é o intervalo de saída da rede [Huang et al., 1996].

A arquitetura da rede era formada por 9 neurônios na camada de entrada, 1 neurônio na camada de saída e 12 neurônios na camada escondida (este número foi escolhido depois de vários testes de arquitetura da rede). O coeficiente de correlação, que indica a dispersão, foi de 0.8. Os resultados foram comparados com métodos estatísticos convencionais: Regressão Linear Múltipla e Regressão Não-Linear Múltipla que apresentaram coeficiente de correlação 0.65 e 0.66 respectivamente. No conjunto de teste o coeficiente de correlação foi de 0.85. Os resultados com RNAs, portanto, se mostraram melhores, indicando maior precisão que os métodos convencionais [Huang et al., 1996].

Outro trabalho sobre predição de permeabilidade pode ser encontrado em [Mohaghegh et al., 1995]. Neste projeto foi usada uma Rede Neural *Backpropagation* que adota um treinamento supervisionado. A rede era composta por 3 camadas com 15 neurônios na camada escondida. Foram selecionados dados de perfis e testemunhos de cinco poços. Haviam 151 exemplos de testemunhos disponíveis, 23 dos quais foram separados para teste (dados de testemunhos e perfis combinados) e o restante foi usado para treinamento.

Para treinar a rede foram utilizados dados de profundidade, raios gama, densidade, indução, especificação da subdivisão zonal e permeabilidade medidas a partir de cada exemplo de testemunho. O coeficiente de correlação alcançado foi de 0.963, onde 1 é a correlação perfeita [Mohaghegh et al., 1995].

Em [Aminian et al., 2000] é apresentado o uso de Redes Neurais na predição de permeabilidade a partir de perfis de poços para auxiliar a simulação do desempenho da recuperação secundária.

A porosidade é uma das propriedades fundamentais das rochas reservatório. Geralmente, a porosidade é obtida a partir de perfis e testemunhos. No entanto, o processo de teste-

munhagem é muito caro. Em [Al-Qahtani, 2000] pode ser encontrado uma aplicação de Redes Neurais Artificiais para a predição da distribuição de porosidade.

Em [He et al., 2001; Yang et al., 2000] pode ser encontrado uma abordagem de Redes Neurais para predizer o desempenho de produção de poços de óleo baseados na variação espacial e séries de tempo, ou seja, a história dos dados de produção dos poços mais próximos e dados de produções anteriores dos poços.

Outra aplicação possível é a geração de perfis sintéticos de Imagens de Ressonância Magnética (MRI - *Magnetic Resonance Imaging*) usando dados de perfis convencionais. MRI são perfis de poços que usam ressonância magnética nuclear para medir precisamente fluido livre, água irredutível e porosidade efetiva [Mohaghegh et al., 1998].

Também é possível aplicar Redes Neurais para predizer o desempenho de um poço após o fraturamento em campos de gás na ausência de dados do reservatório [Mohaghegh et al., 1996] e na identificação de parâmetros que influenciam na reação dos poços de armazenamento de gás em fraturamentos hidráulicos, também na ausência de dados suficientes do reservatório [McVey et al., 1994].

4.4 Sumário

Este capítulo apresentou uma revisão bibliográfica das técnicas de identificação de litofácies, das técnicas de extração de regras de Redes Neurais Artificiais e de outras aplicações de Redes Neurais na indústria do petróleo, temas fortemente relacionados a esta dissertação.

Através desta revisão bibliográfica verificamos que Redes Neurais apresenta-se como uma técnica promissora para a resolução de inúmeros problemas da Indústria do Petróleo, incluindo a identificação automática de litofácies a partir de dados de perfis, tema central desta dissertação.

Adicionalmente, verificamos a existência de vários algoritmos disponíveis para sanar um dos principais problemas com o uso de Redes Neurais como ferramenta de mineração de dados: a dificuldade de se obter explicações acerca do conhecimento adquirido.

O próximo capítulo descreverá o problema de reconhecimento de litofácies de um reservatório de petróleo, apresentará o método proposto neste trabalho e os dados utilizados na resolução do problema.

Capítulo 5

O Problema de Identificação das Litofácies de um Reservatório de Petróleo

Durante a perfuração de um poço de petróleo, as amostras de calha que saem na lama de perfuração são continuamente analisadas, sendo registradas as profundidades associadas a cada tipo de rocha identificada. Após a perfuração são descidas várias ferramentas para medir algumas propriedades da formação, os perfis, conforme explicado na Seção 2.5 do Capítulo 2.

Com os perfis registrados e com base nas observações das amostras de calha, realizadas durante a perfuração, as curvas de perfilagem são interpretadas por um geólogo para saber que tipos de rocha existem naquela formação e a que profundidade. Desta forma pode-se saber se há ou não indícios de hidrocarbonetos naquela formação [Thomas, 2001].

Para predizer o desempenho do reservatório de forma confiável, é necessário fazer uma descrição precisa do mesmo. A testemunhagem é uma das técnicas mais antigas e ainda praticadas para extrair características de um reservatório. No entanto, testemunhar todos os poços em um campo muito grande pode ser economicamente inviável, além disso, o tempo consumido pode ser muito grande. Já os perfis de poços, estão disponíveis para todos os poços.

Devido aos problemas citados acima, torna-se de grande utilidade fazer o reconhecimento automático das litofácies (tipo de formação da rocha) de um reservatório utilizando-se perfis de poços. Este capítulo visa mostrar a importância do reconhecimento automático de litofácies em um reservatório e apresentar um método para resolver este problema. Neste

capítulo também são apresentados os dados utilizados na etapa de experimentação, bem como os critérios adotados para selecioná-los.

5.1 Identificação Automática de Litofácies

A identificação manual de litofácies de um reservatório de petróleo é um processo intensivo que envolve o gasto de uma quantidade considerável de tempo por parte de um especialista experiente. O problema se torna muito mais difícil à medida que aumenta o número de perfis (medidas de determinadas propriedades da formação geológica) simultâneos a serem analisados.

A identificação de litofácies pode ser feita a partir de dados de perfis ou de dados de testemunhos. Os dados de perfis e testemunhos carregam informações diferentes sobre a litologia, por isso a determinação de litofácies a partir destas duas fontes é diferente. As principais técnicas de identificação de litofácies já foram apresentadas na Seção 4.1 do Capítulo 4.

A caracterização de reservatórios de petróleo é uma tarefa muito complexa devido à heterogeneidade dos mesmos. Os reservatórios heterogêneos são conhecidos pelas grandes mudanças em suas propriedades dentro de uma pequena área. Estas mudanças ocorrem, principalmente, devido às idades geológicas distintas, à natureza da rocha e aos ambientes deposicionais. A caracterização de reservatórios tem um papel muito importante na indústria do petróleo, particularmente para o sucesso econômico do gerenciamento e dos métodos de produção [Al-Qahtani, 2000].

Conforme citado na introdução deste capítulo, a testemunhagem é realizada para apenas alguns poços escolhidos estrategicamente, enquanto os perfis estão disponíveis para todos os poços. As informações detalhadas das litofácies de poços testemunhados raramente são extrapoladas e incorporadas quantitativamente durante o estágio de modelagem de um reservatório. Geralmente, nos poços não testemunhados, as modelagens de litofácies são construídas usando as interpretações baseadas em procedimentos de correlação padrão, onde são usados apenas dados limitados de perfilagem de baixa resolução, como por exemplo, raios gama. Isto acontece porque, além da disponibilidade destes perfis, seus comportamentos são muito bem conhecidos [Coll et al., 1999].

A identificação de litofácies a partir de dados de perfis não permite um nível de detalhamento idêntico ao do testemunho. Isto ocorre porque, nos testemunhos, as litofácies são identificadas e classificadas com base em características medidas em pequena escala e em escalas microscópicas [Coll et al., 1999]. No entanto, para este problema, o alto nível de detalhamento dos testemunhos prejudica o processo de aprendizagem. Outros fatores que prejudicam a aprendizagem, mas que dizem respeito aos perfis, são a pouca quantidade de sensores, ruído presente nos dados e maior imprecisão dos perfis em relação aos testemunhos.

Um dos principais problemas da associação dos dados de perfis e testemunhos diz respeito à localização das litofácies presentes no testemunho na profundidade correta nos perfis. Isto ocorre porque, geralmente, há um pequeno deslocamento da profundidade do testemunho com relação à profundidade do perfil. Embora os dados utilizados já estivessem com o deslocamento devidamente acertado, por tratar-se de um processo manual, este alinhamento pode ser impreciso.

O método proposto neste trabalho para resolver o problema do reconhecimento de litofácies é baseado em uma abordagem de Redes Neurais. Os principais motivos que levaram à escolha do uso desta técnica já foram discutidos na Seção 3.1.3 do Capítulo 2. Na próxima seção será mostrado o método proposto neste trabalho.

5.2 Método Proposto

O método proposto neste trabalho para o problema de reconhecimento de litofácies, bem como para melhorar a interpretação dos resultados é composto pelas seguintes etapas: associação dos dados, discretização dos dados, agrupamento de classes, treinamento da Rede Neural, tratamento dos padrões problemáticos, extração de regras e validação. A idéia é que o método seja genérico, no entanto, a associação dos dados precisará ser realizada novamente, as redes precisarão ser retreinadas e os padrões problemáticos tratados, quando utilizando dados de outros campos. Mesmo sem apresentar um resultado compreensível, o método proposto pode ser muito útil como ferramenta de apoio a tomadas de decisões. A Figura 5.1 mostra a arquitetura geral descrita. Implementações foram desenvolvidas e experimentos foram realizados envolvendo todos as etapas do método, exceto a etapa de extração

de regras cuja implementação fugiria ao escopo desta dissertação. O Capítulo 6 detalha melhor a parte experimental do trabalho. A seguir cada etapa do método proposto será discutida brevemente.

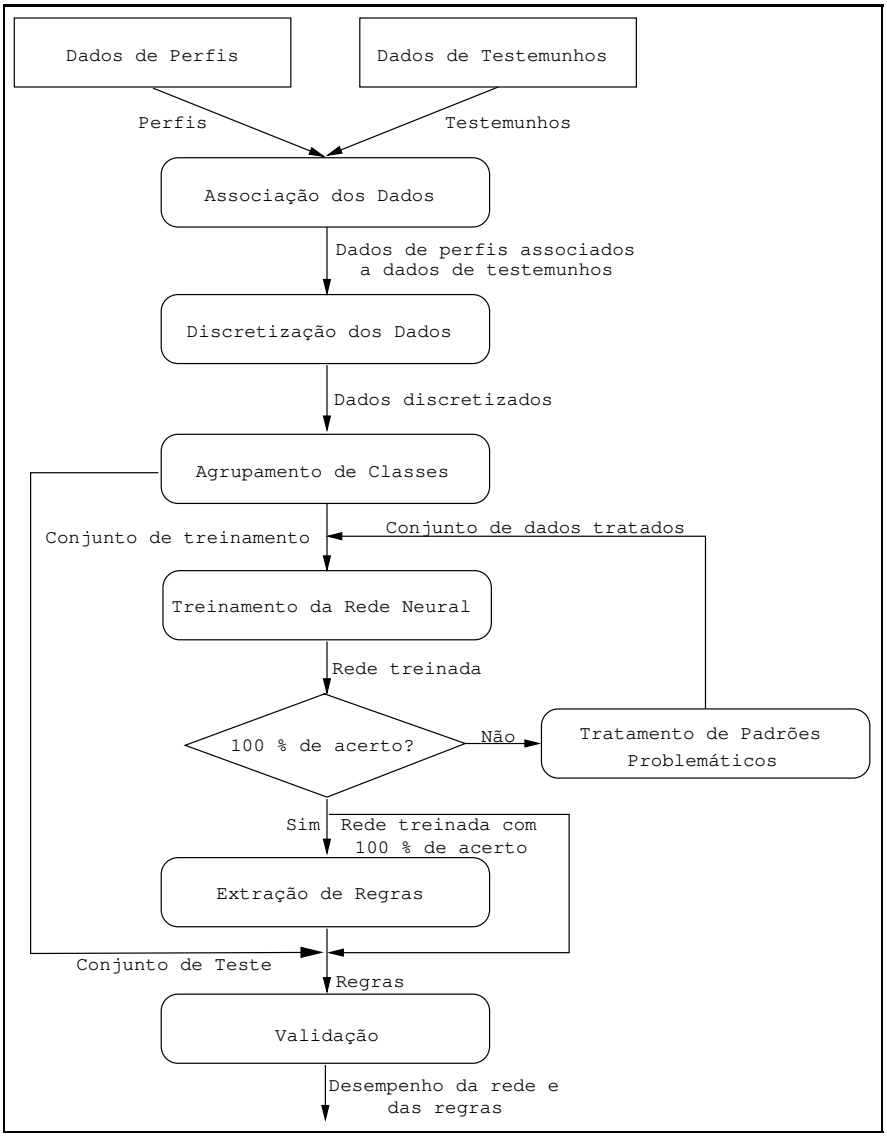

Figura 5.1: Arquitetura geral do método proposto.

5.2.1 Associação dos Dados

Na etapa de associação dos dados, os dados de perfis e testemunhos de cada poço selecionado foram associados de acordo com a profundidade. Geralmente, há um pequeno deslocamento da profundidade do testemunho com relação à profundidade do perfil, pois a resolução das ferramentas não é exatamente igual. Para associar os perfis aos testemunhos através da

profundidade é necessário comparar as curvas de radioatividade de ambos e realizar o deslocamento necessário. Os detalhes de como esta associação foi feita se encontram no Capítulo 6.

5.2.2 Discretização dos Dados

A discretização dos dados é necessária para facilitar a tarefa de identificação de padrões nos dados. Uma Rede Neural requer que todas as suas entradas sejam numéricas, portanto, os atributos nominais[1] devem ser transformados para uma representação numérica. Algumas vezes os atributos numéricos também precisam ser discretizados para facilitar as etapas seguintes. Na etapa de discretização dos dados, foram aplicadas transformações de forma que os dados ficassem num mesmo intervalo.

5.2.3 Agrupamento de Classes

A etapa de agrupamento de classes foi incluída na arquitetura após ter sido constatado que a classificação das litofácies utilizadas estava muito detalhada e, por consequência, estava prejudicando a tarefa de reconhecimento. Um determinado tipo de rocha pode ter várias subdivisões de acordo com suas propriedades. A classificação pode ser considerada muito detalhada quando, dentro de um pequeno intervalo de profundidade, existem várias subdivisões de rochas. Este nível de detalhamento geralmente só é alcançado em laboratório com a utilização de equipamentos específicos. Algumas litofácies podem ser agrupadas sem prejudicar a consistência das classes, pois pertencem a um mesmo tipo de rocha e o que as diferenciam são pequenas variações em suas propriedades, como por exemplo sua granularidade.

5.2.4 Treinamento das Redes Neurais

Para treinar as redes foi utilizado um simulador de Redes Neurais, o SNNS (*Stuttgart Neural Network Simulator*) [SNNS,]. A arquitetura utilizada nos experimentos foi *Multilayer Perceptron* com algoritmo de treinamento *Backpropagation* (ver Seção 3.1.3 do Capítulo 2).

[1] Atributos alfa-numéricos.

Seja um padrão de entrada p, $p = 1, 2, ..., P$, os valores das unidades de saída da rede, S_{ip}, e os valores de ativação das unidades escondidas, H_{jp} são:

$$S_{ip} = \sigma\left(\sum_{j=1}^{J} v_{ij} H_{jp}\right) \tag{5.1}$$

$$H_{jp} = \sigma(W_j X_p) = \sigma\left(\sum_{k=1}^{K} w_{jk} x_{kp}\right) \tag{5.2}$$

onde $x_{kp} \in [0, 1]$ é o valor da unidade de entrada k dado um padrão p, w_{jk} é o peso da conexão da unidade de entrada k para a unidade escondida j, v_{ij} é o peso da conexão da unidade escondida j para a unidade de saída i e $\sigma(\xi)$ é a função sigmóide $1/(1 + e^{-\xi})$. J e K são os números de unidades escondidas e de entrada, respectivamente. Cada padrão x_p pertence a uma das C classes possíveis: $C_1, C_2, ..., C_c$. O valor esperado para o padrão p na unidade de saída i é chamado t_{ip}. O número de unidades de saída na rede é igual ao número de classes. Se o padrão p pertence a classe c, então $t_{cp} = 1$ e $t_{ip} = 0, \forall i \neq c$.

Várias Redes Neurais foram treinadas com os dados discretizados e agrupados. Após este treinamento, verificou-se que as redes não atingiam a taxa de aprendizagem desejada. Observou-se que as redes não aprendiam padrões que estavam em zonas de transições ou que pertenciam a uma camada muito fina em meio a outro tipo de rocha. Para resolver este problema, foi acrescentada a etapa de tratamento de padrões problemáticos. Caso a rede não atingisse a taxa de aprendizagem esperada, os padrões eram tratados e a rede treinada novamente. Quando os conjuntos não possuiam mais padrões problemáticos, as redes alcançavam uma taxa de 100 % de acerto no conjunto de treinamento. Cabe lembrar que, assim como ocorre com regras de classificação perfeitas, redes treinadas com 100 % de acerto na fase de treinamento nem sempre são confiáveis (por exemplo, quando o tamanho do conjunto de treinamento não é significativo).

5.2.5 Tratamento de Padrões Problemáticos

Nesta etapa, os padrões não aprendidos pela rede, foram analisados manualmente e constatou-se que eles estavam em zonas de transição de litofácies ou pertenciam a camadas muito finas, conforme citado na seção anterior. Estes padrões ou migraram para classes de camadas adjacentes (re-classificação), ou foram removidos. Caso a camada adjacente não es-

tivesse no conjunto de treinamento, o padrão era eliminado do conjunto. Este tratamento de padrões problemáticos não é ideal, pois algumas camadas muito finas são importantes para a identificação de reservatórios. Caso a existência destas camadas seja omitida devido a uma re-classificação ou eliminação do conjunto de treinamento corre-se o risco da caracterização do reservatório ficar incoerente.

5.2.6 Extração de Regras

Os resultados otimizados obtidos do treinamento das Redes Neurais, devem ser apresentados como entrada para um algoritmo de extração de regras, de forma que o conhecimento adquirido seja humanamente mais compreensível. O algoritmo de extração de regras de Redes Neurais escolhido para a aplicação deste trabalho foi o desenvolvido por Setiono e Leow [Setiono and Leow, 2000; Setiono and Leow, 1999], descrito no Apêndice A, com algumas alterações. Este algoritmo foi selecionado devido a sua simplicidade em relação aos outros. Geralmente, as unidades e as conexões redundantes e irrelevantes de uma Rede Neural treinada são removidos por um algoritmo de poda antes que as regras sejam extraídas. Estes algoritmos requerem que a rede seja retreinada, consumindo muito tempo de processamento. O algoritmo utilizado neste trabalho não necessita poda e portanto nenhum retreinamento é necessário.

A diferença entre o algoritmo FERNN e o algoritmo estudado neste trabalho é que, no primeiro é utilizada uma função de erro *cross-entropy* aumentada por uma função de penalização, de forma que as conexões irrelevantes tenham pesos baixos, enquanto que no segundo foi utilizada a função de erro padrão do *Backpropagation* [Beale and Jackson, 1990].

Os dados utilizados nos experimentos foram selecionados de uma suite de dados do Campo Escola de Namorado, a qual será detalhada na próxima seção.

5.2.7 Validação

A etapa de validação foi dividida em duas fases: validação da Rede Neural e validação das regras extraídas da rede treinada. Como o algoritmo de extração de regras não foi implementado, a validação das regras extraídas não foi realizada. O principal objetivo da validação das regras é verificar se as mesmas classificam padrões desconhecidos sem diminuir a taxa

de acerto alcançada pela rede.

A validação da Rede Neural foi realizada com o objetivo de verificar se a rede treinada era capaz de reconhecer padrões desconhecidos. Para isto, após o treinamento, um conjunto de dados era apresentado à rede e a taxa de acerto era calculada.

5.3 Dados do Campo Escola de Namorado

A Agência Nacional do Petróleo (ANP) disponibilizou dados para pesquisa do Campo Escola de Namorado - Bacia de Campos, RJ. Estes dados são compostos de:

- Dados sísmicos 2D e 3D, em formato SEGY [Segy, 2002; SEG, 2002], conforme abaixo:
  - Linhas Sísmicas 2D Migradas;
  - Programa Sísmico 3D Migrado.
- Arquivos em formato LAS com a suite básica de curvas (GR, ILD, RHOB, NPHI e DT) de 56 poços;
- Mapabase e arcabouço de seções estruturais e estratégias
- Outros Dados: Aplicativo de visualização e busca ("DocReader") para acesso ao arquivo ("Namorado.docpro") contendo:
  - Descrições dos testemunhos - Formato Anasete - de 19 poços;
  - Dados Petrofísicos:
    - 594 medidas de porosidade e permeabilidade de 15 poços;
    - 10 medidas de pressão capilar e 12 medidas de permeabilidade relativa de 2 poços;
    - 10 medidas de compressibilidade de rocha de 3 poços;
  - Estudos de propriedades de fluidos (PVT): 8 análises de 7 poços;
  - Histórico de vazões de óleo, gás e água;

- Mapas (posicionamento sísmico 2D e 3D, posicionamento de poços) e Listagens de dados do Campo de Namorado (Sísmica, Poços, Vértices do Campo);
- Legislação e regulação concernentes aos dados.

5.3.1 Dados de Perfis

Os dados de perfis se encontram no formato LAS, que é composto basicamente por um cabeçalho contendo informações sobre o poço, sobre os perfis medidos e por colunas numéricas, em que cada coluna representa um perfil com exceção da primeira que indica a profundidade em que a propriedade foi medida, conforme ilustrado na Figura 5.2. A Figura 5.3 ilustra um exemplo das curvas produzidas a partir dos dados de perfis.

As informações contidas no cabeçalho, referentes ao poço, são as seguintes: profundidade do início da perfilagem (caso o poço seja marítimo, a lâmina d'água não entra no cálculo da profundidade); profundidade final da perfilagem, intervalo de medição do perfil (em metros), representação para valores nulos, nome da companhia, nome do poço, nome do campo, localização, estado, nome da companhia contratada para realizar a perfilagem, data em que a perfilagem foi realizada e código API. Em seguida, o cabeçalho contém o nome das colunas de dados, ou seja, a profundidade e os perfis que foram medidos.

5.3.2 Dados de Testemunhos

Os dados de testemunho estão no formato ANASETE, que consiste de uma representação gráfica contendo as informações retiradas do testemunho. A Figura 5.4 apresenta um exemplo dos dados de testemunho, onde a coluna 1 representa a profundidade em relação ao perfil, a coluna 2 representa a profundidade em relação à ferramenta de testemunhagem, a coluna 3 representa o número da caixa de armazenamento do testemunho, a coluna 4 representa o tamanho das caixas em que o testemunho foi armazenado, a coluna 5 representa a granulometria, a coluna 6 representa as estruturas das rochas e, finalmente, a coluna 7 representa as litofácies encontradas.

Existem 28 tipos de litofácies nos testemunhos dos poços selecionados. A Tabela 5.1 apresenta os nomes das litofácies utilizadas nos experimentos. Os principais tipos de rocha no contexto da exploração de óleo são os arenitos e os folhelhos. Os arenitos são rochas

```
~VERSION INFORMATION
 VERS.           2.0: CWLS Log ASCII Standard - Version 2.0
 WRAP.            NO: One Line per Depth Step
~WELL INFORMATION
#MNEM.UNIT              DATA            DESCRIPTION OF MNEMONIC
#----------            ----            -----------------------
 STRT.M          2975.0000              :Start Depth
 STOP.M          3200.0000              :Stop  Depth
 STEP.M             0.2000              :Step
 NULL.             -99999.0             :Null Value
 COMP  .   PETROLEO BRASILEIRO S/A      :Company
 WELL  .   3NA  0002  RJS               :Well
 FLD   .   NAMORADO                     :Field
 LOC   .                                :Location
 STAT  .   RIO DE JANEIRO               :State
 SRVC  .                                :Service Company
 DATE  .                                :Log Date
 API   .   742810010100                 :API Code
~CURVE INFORMATION
 DEPT.M                                 :Measured Depth
 DT.                                    :01
 GR.                                    :02
 ILD.                                   :03
 NPHI.                                  :04
 RHOB.                                  :05
~ASCII LOG DATA
  2975.000    91.2695    66.4531    1.8425    21.7996    2.4617
  2975.200    90.9399    68.4648    1.7627    21.8086    2.4266
  2975.400    90.3281    69.0938    1.7114    23.1367    2.4270
  2975.600    87.7031    67.7969    1.6621    24.6211    2.4411
  2975.800    85.2031    67.1992    1.5940    25.5234    2.4502
  2976.000    86.5234    67.9570    1.5305    26.3086    2.4579
  2976.200    91.1875    69.0273    1.4912    26.2773    2.4529
  2976.400    94.4570    69.8750    1.4514    25.7148    2.4197
  2976.600    95.6133    70.5117    1.4237    25.1562    2.3845
  2976.800    97.0078    70.3398    1.4192    24.8650    2.3754
  2977.000    97.0391    68.6406    1.4285    25.0002    2.3744
  2977.200    96.2578    66.5039    1.4381    25.1091    2.3668
  2977.400    96.0483    65.2070    1.4400    25.0427    2.3782
  2977.600    95.9258    64.8555    1.4288    25.1250    2.3969
For Help, press F1
```

Figura 5.2: Exemplo de dados de perfis no formato LAS.

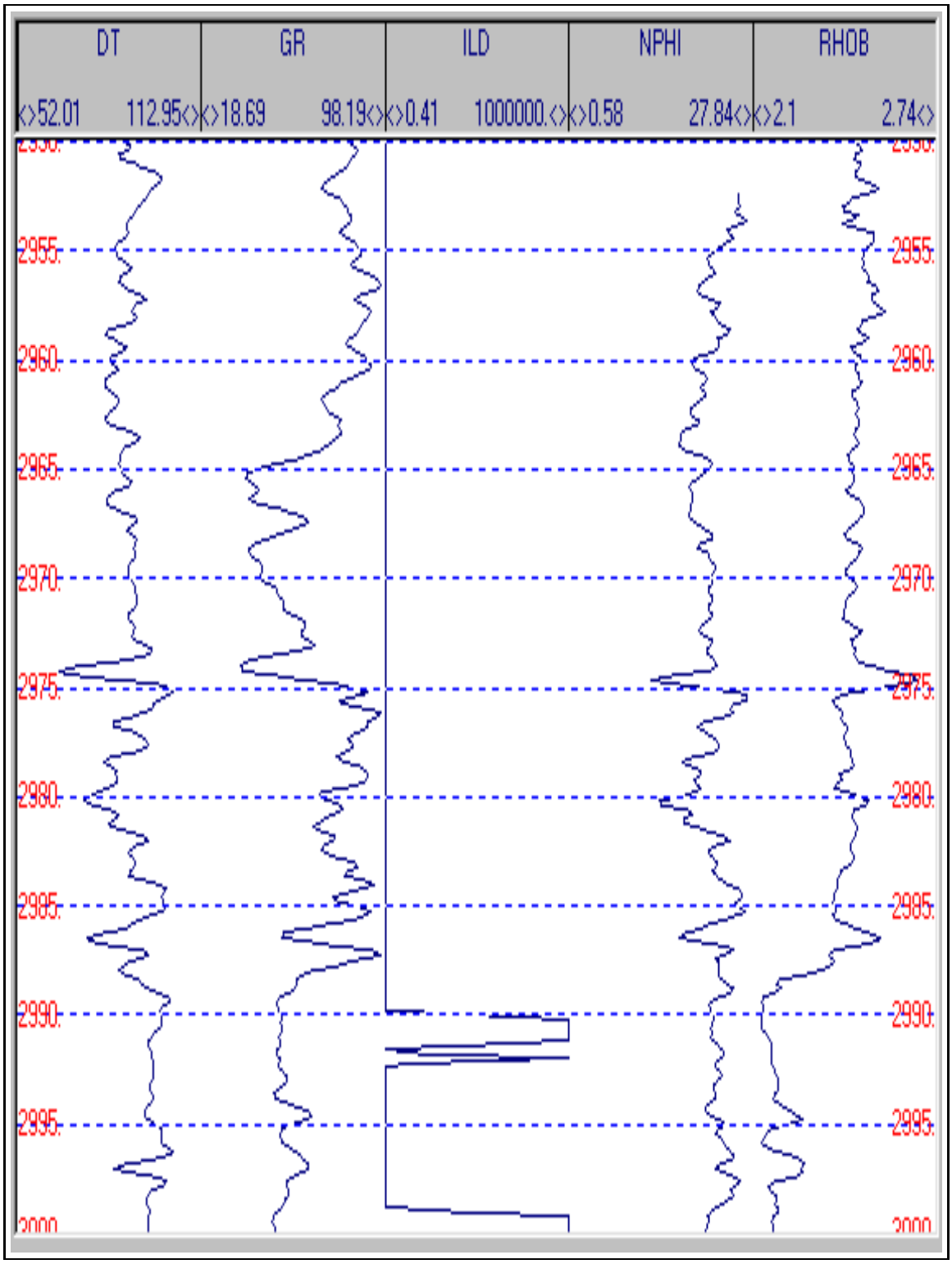

Figura 5.3: Curvas produzidas a partir dos dados de perfis.

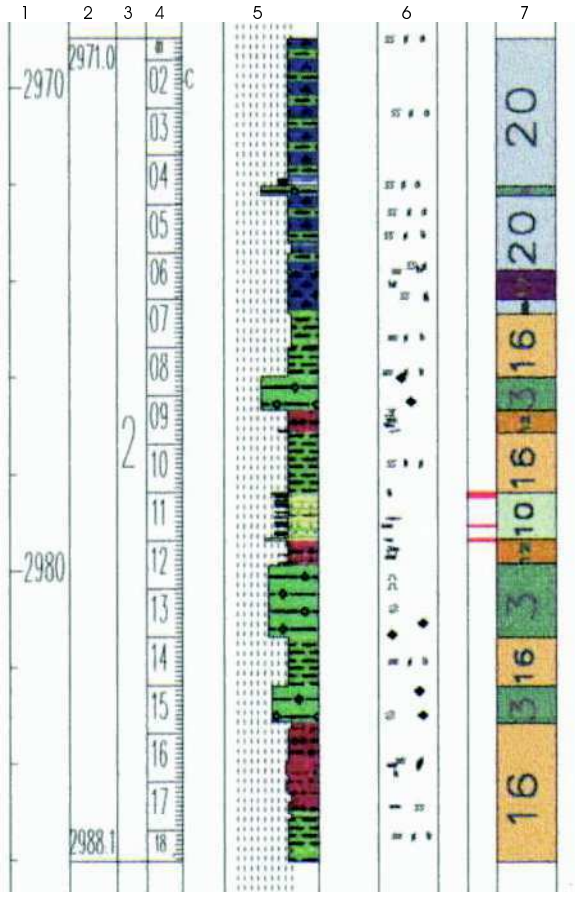

Figura 5.4: Segmento de um testemunho do Campo de Namorado.

que possuem alta porosidade e permeabilidade, ficando evidente que estes tipos de rocha podem constituir os reservatórios petrolíferos. Os folhelhos são rochas que oferecem as condições físico-químicas necessárias para a geração do óleo. Na maioria das vezes as rochas geradoras de óleo são folhelhos. Este tipo de rocha também constitui as rochas selantes, rochas que garantem o aprisionamento do óleo devido a sua baixa permeabilidade, conforme discutido no Capítulo 2. Excepcionalmente, os folhelhos podem reter petróleo e os arenitos não. No caso da Tabela 5.1, as litofácies 6, 7 e 8, por exemplo, podem apresentar indícios de hidrocarbonetos, observados a partir dos testemunhos. Como exemplo dos folhelhos na Tabela 5.1 temos as litofácies 14 e 20.

Litofácie	Nome
1	Interlaminado Lamoso Deformado
2	Conglomerado e Brechas Carbonáticas
3	Diamictito Areno Lamoso
4	Conglomerados Residuais
5	Arenito com Intraclastos Argilosos
6	Arenito Grosso Amalgamado
7	Arenito Médio Laminado
8	Arenito Médio Maciço Gradado
9	Arenito Médio Cimentado
10	Arenito / Folhelho Interestratificado
11	Arenito / Folhelho Finamente Interestratificado
12	Siltito Argiloso Estratificado
13	Interlaminado Siltito Argiloso e Marga
14	Folhelho Radioativo
15	Interlaminado Arenoso Bioturbado
16	Interlaminado Siltito e Folhelho Bioturbado
17	Marga Bioturbada
18	Ritmito
19	Arenito Glauconítico
20	Folhelho com Níveis de Marga Bioturbados
21	Arenito Cimentado com Intraclastos
22	Siltito Argiloso / Arenito Deformado
23	Arenito Médio / Fino Laminado Cimentado
24	Interestratificado Siltito / Folhelho Intensamente Bioturbados
25	Folhelho Carbonoso
26	Arenito Maciço muito fino
27	Siltito Areno-Argiloso
28	Interlaminado Siltito Folhelho

Tabela 5.1: Litofácies disponíveis para realização dos experimentos.

5.3.3 Dados Selecionados

Com o intuito de resolver o problema de reconhecimento de litofácies, foram selecionados os arquivos de perfis e descrição dos testemunhos do Campo Escola de Namorado. Nem todos os poços que possuiam dados de perfis, possuiam dados de testemunho, mas todos os poços que possuiam dados de testemunho, possuiam dados de perfis. Para alguns poços com dados de testemunho, os dados de perfis estavam incompletos, portanto, foram selecionados apenas os poços que, além do testemunho, possuiam todos os perfis. No final restaram apenas 9 poços com descrições dos testemunhos, além dos seguintes perfis: Raios Gama (GR), Sônico (DT), Indução (ILD), Densidade (RHOB) e Porosidade Neutrônica (NPHI), conforme descreve a Tabela 5.2.

POÇO	**INTERVALO DE PROFUNDIDADE(m)**
3NA 0001A RJS	2950,0 - 3200,0
3NA 0002 RJS	2975,0 - 3200,0
3NA 0004 RJS	2950,0 - 3150,0
7NA 0007 RJS	3025,0 - 3275,0
7NA 0011A RJS	3000,0 - 3200,0
7NA 0012 RJS	2970,0 - 3175,0
7NA 0037D RJS	3170,0 - 3400,0
4RJS 0042 RJ	3000,0 - 3215,0
4RJS 0234 RJ	3150,0 - 3352,2

Tabela 5.2: Poços selecionados.

5.4 Sumário

Neste capítulo foi discutida a importância da identificação automática de litofácies. Em seguida foi apresentado o método proposto neste trabalho para identificar litofácies de forma automática através de Redes Neurais Artificiais, bem como, para melhorar a apresentação dos resultados, com uma breve descrição de cada etapa do método. Também foi apresentado neste capítulo a origem dos dados utilizados para experimentação e quais destes dados foram

selecionados.

No próximo capítulo será apresentada toda a sequência de experimentações realizadas, com seus respectivos resultados. Será apresentada também uma discussão dos resultados obtidos.

Capítulo 6

Experimentos e Resultados

Neste capítulo serão apresentados experimentos e resultados relacionados ao método proposto para resolver o problema de identificação de litofácies, discutido no capítulo anterior. Primeiramente, serão apresentados experimentos realizados com dados do perfil Potencial Espontâneo, obtidos em um curso de Geologia do Petróleo. Nesta fase, é utilizado um método de janelamento o qual consiste em apresentar vários padrões de uma só vez à rede, além de um treinamento padrão (apresentação de apenas um padrão à rede). Em seguida, serão apresentados os primeiros experimentos com os dados do Campo de Namorado. Após estes experimentos detectou-se diversos problemas com relação aos dados. Para resolver estes problemas, uma nova abordagem de pré-processamento é inserida, o agrupamento de litofácies, ou eletrofácies. Os experimentos e resultados desta fase são apresentados na última seção deste capítulo.

6.1 Experimentos com Perfis de Potencial Espontâneo

Com o objetivo de validar o método proposto, iniciamos a etapa de experimentação utilizando dados de dois perfis que foram analisados manualmente com o auxílio de um geólogo, sem utilizar informações de testemunhos. Os dados utilizados nestes experimentos foram do perfil Potencial Espontâneo (SP) - registro da diferença de potencial entre um eletrodo móvel descido dentro do poço e outro fixo na superfície - e profundidade de dois poços de localização desconhecida. A Figura 6.1 mostra um segmento do perfil utilizado nestes experimentos. O perfil estava impresso em papel milimetrado. Os dados foram digitalizados

levando em consideração a reta em negrito na figura (valores à direita foram considerados positivos e valores à esquerda negativo). Os valores da curva correspondiam à distância, em centímetros, até a reta. Cada divisão do papel era vista como uma profundidade. Os dados foram discretizados conforme a Equação 6.1.

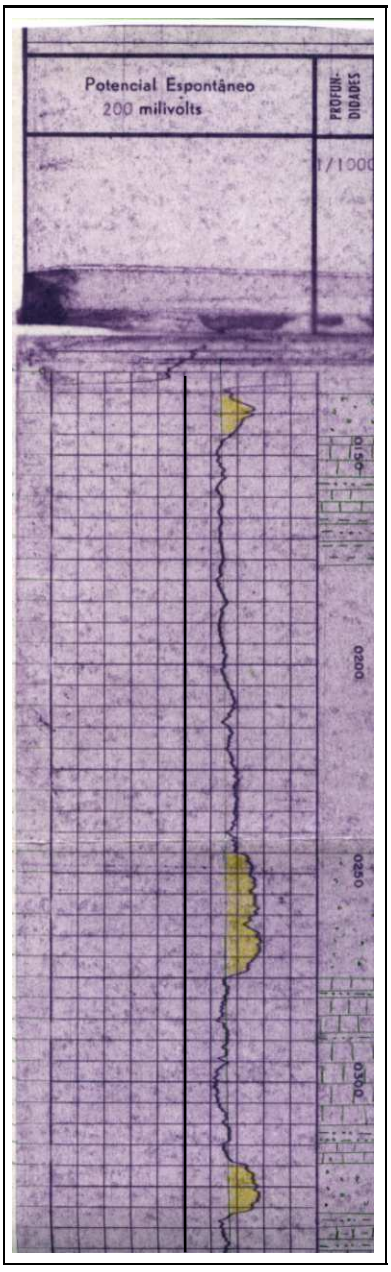

Figura 6.1: Exemplo de um perfil de Potencial Espontâneo.

$$ValorNormalizadodoPerfil = \frac{ValorPerfil - menorPerfil}{maiorPerfil - menorPerfil} \tag{6.1}$$

em que $menorPerfil$ e $maiorPerfil$ representam o menor e o maior valor do perfil em toda a base de dados, respectivamente.

Segundo [White et al., 1995], não é possível identificar as litofácies de um reservatório usando apenas os perfis dos poços, portanto as inclinações das curvas nos gráficos dos perfis com relação ao ponto anterior e posterior (ver Figura 5.3 na Seção 5.3.1 do Capítulo 5) foram adicionadas aos dados. Para calcular as inclinações anterior e posterior das curvas foi utilizada a derivada discreta em cada ponto, conforme as Equações 6.2 e 6.3, respectivamente.

$$InclinacaoAnterior_i = \frac{ValorPerfil_i - ValorPerfil_{i-1}}{Profundidade_i - Profundidade_{i-1}} \tag{6.2}$$

$$InclinacaoPosterior_i = \frac{ValorPerfil_i - ValorPerfil_{i+1}}{Profundidade_i - Profundidade_{i+1}} \tag{6.3}$$

em que $ValorPerfil_i$ e $Profundidade_i$ representam o $i-$ésimo valor do perfil e o i-ésimo valor da profundidade, respectivamente.

Para normalizar os valores das inclinações no intervalo [0, 1], foi utilizada a Equação 6.4.

$$ValorNormalizadodaInclinacao_i = \frac{ValorInclinacao_i - menorInclinacao}{maiorInclinacao - menorInclinacao} \tag{6.4}$$

em que $ValorInclinacao_i$ representa o i-ésimo valor da inclinação anterior/posterior, $menorInclinacao$ e $maiorInclinacao$ representam o menor e o maior valor da inclinacao anterior/posterior de um perfil em toda a base de dados, respectivamente.

Método da Janela de Treinamento

Primeiramente, metade dos dados de um poço foi utilizada para treinamento e a outra metade para teste. Nestes experimentos foi apresentada uma "janela" do perfil à rede. Uma janela corresponde a uma sequência de n padrões a partir de uma determinada profundidade. Ao utilizar a técnica da janela, a Rede Neural será alimentada com um segmento do gráfico (perfil), ou seja, vários pontos do perfil serão apresentados simultaneamente à rede, o que pode colaborar na tarefa de reconhecimento das litofácies.

Várias redes foram treinadas variando o tamanho da janela. A Tabela 6.1 mostra um resumo dos experimentos realizados (utilizando apenas dados de um poço). A coluna N^o **de**

Neurônios indica o número de neurônios da camada escondida (o que corresponde exatamente ao número de neurônios na camada de entrada), a coluna **Tamanho da Janela** indica os valores associados a uma sequência de profundidades apresentadas à rede de uma só vez e a coluna **Taxa de Acerto** indica a taxa de acerto média para todo o conjunto de teste. Nesta etapa do trabalho não havia a preocupação em decidir uma arquitetura ideal para as Redes Neurais, por isso optou-se por utilizar um número idêntico de neurônios nas duas camadas iniciais.

N^o **de Neurônios**	**Tamanho da Janela**	**Taxa de Acerto**
20	5	34,79 %
24	6	34,21 %
28	7	33,74 %
32	8	31,72 %
36	9	32,37 %
40	10	28,90 %

Tabela 6.1: Resumo dos experimentos realizados com o perfil SP de apenas um poço.

Nestes experimentos, a saída possuia dois neurônios, um representando a classe dos arenitos e o outro representando tudo que não era arenito, codificados na saída da rede com 0 1 e 1 0, respectivamente. Na Tabela 6.1 pode-se observar que conforme aumenta o tamanho da janela, a taxa de acerto diminui. Isto nos leva a acreditar que a não utilização da janela pode produzir uma taxa de acerto maior.

Reta-base

Posteriormente, na tentativa de melhorar os resultados da Tabela 6.1, foram realizados experimentos incluindo nos dados uma coluna adicional indicando se a curva do perfil estava à direita ou à esquerda da reta-base. A reta-base é uma reta traçada manualmente sobre os pontos que não oscilam. Na análise manual de perfis, a reta-base é muito importante para a determinação da variação do perfil em relação à medidas homogêneas. Desta forma, juntamente com outras informações (como a resistividade), é possível saber quais os intervalos de profundidade possuem indícios de hidrocarboneto. A Tabela 6.2 mostra os resultados destes

experimentos. Como foi inserida uma nova coluna nos dados o número de neurônios das camadas de entrada e escondida aumentou.

N^o **de Neurônios**	**Tamanho da Janela**	**Taxa de Acerto**
25	5	55,62 %
30	6	50,88 %
35	7	49,09 %
40	8	46,77 %
45	9	43,96 %
50	10	39,12 %

Tabela 6.2: Resumo dos experimentos realizados com o perfil SP de apenas um poço incluindo a indicação da posição do perfil em relação à reta-base.

Na Tabela 6.2 pode-se observar o mesmo comportamento da Tabela 6, o que indica novamente que a eliminação da janela pode produzir melhores taxas de acerto. No entanto, neste caso, obteve-se um aumento significativo das taxas de acerto em todos os tamanhos de janela. Percebe-se, portanto, que a inclusão da coluna adicional aos dados de treinamento e teste pode influenciar positivamente o processo de reconhecimento.

Em seguida, investigamos se a remoção da janela teria algum efeito nas taxas de reconhecimento obtidas anteriormente. Da mesma forma anterior, a camada escondida era composta pelo mesmo número de neurônios da camada de entrada. Treinando uma rede com 4 neurônios na camada escondida sem a técnica da janela e sem a coluna indicando a posição da curva em relação à reta-base, a taxa de acerto no conjunto de teste foi de apenas 27 %. No entanto, ao incluir a coluna indicando a posição do perfil com relação à reta-base a taxa de acerto no conjunto de teste aumentou para 73 %.

Ao contrário do previsto, eliminando a janela, a taxa de acerto diminuiu quando a coluna indicando a posição do perfil em relação à reta-base não estava presente. No entanto, quando a coluna adicional estava presente a taxa de acerto aumentou. Isto mostra que para dados com apenas um perfil, a presença da coluna indicando a posição do perfil em relação à reta-base pode aumentar consideravelmente a taxa de acerto, tanto no método com janela como, principalmente, no método sem janela.

Incrementando os Conjuntos de Treinamento e Teste

Posteriormente, utilizamos dados de dois poços diferentes, um para treinamento e o outro para teste, cada um contendo 200 padrões. Utilizamos primeiro o método da janela para um conjunto de dados sem a coluna indicando a posição da curva em relação à reta-base. A Tabela 6.3 mostra os resultados destes experimentos. Comparando esta tabela com a Tabela 6.1, pode-se observar que o incremento no conjunto de treinamento e teste causou o aumento das taxas de acerto. Isto mostra que o acréscimo de dados de novos poços pode melhorar consideravelmente o desempenho das Redes Neurais na tarefa de identificação de litofácies.

N^o **de Neurônios**	**Tamanho da Janela**	**Taxa de Acerto**
20	5	68,06 %
24	6	66,49 %
28	7	68,92 %
32	8	69,82 %
36	9	69,33 %
40	10	70,73 %

Tabela 6.3: Resumo dos experimentos realizados com o perfil SP de dois poços.

Em seguida, apresentamos à rede o conjunto com a coluna indicando a posição da curva em relação à reta-base, também com o método da janela. Os resultados são mostrados na Tabela 6.4. Comparando esta tabela com a Tabela 6.2, confirma-se que o acréscimo de dados de novos poços melhora a taxa de acerto na identificação de litofácies. Ao comparar as Tabelas 6.3 e 6.4 confirma-se a idéia de que o uso da coluna indicando a posição da curva em relação à reta-base melhora os resultados de classificação.

Por fim, o último experimento com os dados do perfil Potencial Espontâneo foi a remoção da janela, utilizando dados de um poço para treinamento e do outro para teste, sem e com a coluna indicando a posição da curva em relação à reta-base. Nestes experimentos, a rede era formada por 4 neurônios na camada de entrada e 4 neurônios na camada escondida para o primeiro caso. No segundo caso, a rede era formada por 5 neurônios na camada de entrada e 5 neurônios na camada escondida. Para cada caso, as taxas de acerto no conjunto de testes foram de 71,5 % e 89 %, respectivamente. A eliminação da janela causou o aumento da

N^o **de Neurônios**	**Tamanho da Janela**	**Taxa de Acerto**
25	5	88,57 %
30	6	89,57 %
35	7	87,85 %
40	8	88,02 %
45	9	89,70 %
50	10	87,59 %

Tabela 6.4: Resumo dos experimentos realizados com o perfil SP de dois poços incluindo a indicação da posição do perfil em relação à reta-base.

taxa de acerto para o conjunto sem a coluna de indicação da posição da curva em relação à reta-base. No entanto, para o conjunto com a coluna adicional a taxa de acerto não obteve aumento.

Tanto nos experimentos em que os dados de treinamento e teste pertenciam ao mesmo poço, quanto nos experimentos em que os dados de treinamento e teste pertenciam a poços distintos, os resultados dos conjuntos com a coluna indicando a posição da curva em relação à reta-base se mostraram melhores. Isto indica que a informação da reta-base pode ser muito útil no aumento das taxas de reconhecimento das Redes Neurais. Entretanto, introduz-se um elemento com processamento manual ao método, já que não existe ainda uma forma automática para determinação da reta-base. Este é o motivo pelo qual a reta-base não foi utilizada nos experimentos com dados reais apresentados a seguir.

6.2 Experimentos Iniciais com Dados do Campo Escola de Namorado

Conforme discutido na Seção 5.2 do Capítulo 5, o primeiro passo da etapa de preparação dos dados foi associar os testemunhos aos perfis de acordo com a profundidade. Os dados utilizados já estavam com o deslocamento de profundidade dos testemunhos e perfis devidamente acertado. Para realizar esta associação, um novo conjunto de dados foi criado contendo os perfis dos 9 poços selecionados (sem o cabeçalho) e as litofácies correspondentes a cada pro-

fundidade, obtidas manualmente a partir dos testemunhos. Cinco poços foram usados para treinamento e quatro selecionados para teste.

Embora todos os valores dos perfis fossem numéricos, foi necessário fazer uma discretização dos mesmos para que eles ficassem padronizados, pois cada perfil estava em um intervalo diferente. Os valores foram normalizados para o intervalo [0, 1]. Para colocar os valores dos perfis neste intervalo, estes foram normalizados da mesma forma que antes (ver Equação 6.1). As inclinações das curvas nos gráficos com relação aos pontos anterior e posterior também foram calculados e normalizados da mesma forma anterior (ver Equações 6.2, 6.3 e 6.4).

As litofácies foram discretizadas de forma que a saída da Rede Neural fosse binária. Cada número j representando um tipo de litofácie foi convertido para um número binário em que todas as posições são nulas, exceto a j-ésima posição cujo valor é 1. Por exemplo, a 20^a litofácie foi representada pelo código binário: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0.

Como foi utilizado um simulador de Redes Neurais, o SNNS (Stuttgart Neural Network Simulator) [SNNS,], para a construção da arquitetura neural, os conjuntos de treinamento e teste foram transformados para um formato especial (.pat), conforme ilustrado na Figura 6.2. As duas primeiras linhas deste formato contém a versão do SNNS, a data e a hora em que o arquivo foi gerado. Em seguida, compondo o cabeçalho juntamente com as linhas anteriores, existem três linhas indicando o número de padrões do conjunto, o número de neurônios de entrada e o número de neurônios de saída. Logo abaixo do cabeçalho começam os dados propriamente ditos. As linhas que contém o símbolo # no início são comentários. O conjunto é intercalado por padrões (entrada da rede) e saídas desejadas (saída da rede).

Nos experimentos iniciais, várias Redes Neurais foram construídas para explorar o espaço de combinações de perfis em busca de uma combinação que produzisse os melhores resultados. Num estágio inicial dos experimentos não utilizamos todas as possíveis combinações, mas apenas um subconjunto. Em cada rede, os parâmetros de entrada foram modificados da seguinte forma: uma rede foi treinada com parâmetros referentes aos perfis ILD e RHOB, outra com GR, ILD e RHOB e assim por diante. Para cada grupo de perfis várias arquiteturas foram criadas, com número diferente de neurônios na camada escondida, com o intuito de verificar a melhor arquitetura para aquele grupo. A Tabela 6.5 mostra um re-

```
SNNS pattern definition file V1.4
generated at Mon Jun 24 18:01:02 2002

No. of patterns     : 509
No. of input units  : 16
No. of output units : 2

# Entrada  1
0.012684 0.410526 0.179966 0.187607 0.131080 0.464557 0.324281 0.001365 0.515344 0.515514
0.607401 0.541102 0.549132 0.777049 0.610184 0.537313
# Saida 1
 0 1

# Entrada  2
0.013191 0.259335 0.258065 0.179966 0.114599 0.535642 0.464557 0.001394 0.515208 0.515344
0.617023 0.464010 0.541102 0.824220 0.639816 0.610184
# Saida 2
 0 1

# Entrada  3
0.013699 0.151074 0.539898 0.258065 0.120904 0.812900 0.535642 0.001307 0.515171 0.515208
0.580122 0.283608 0.464010 0.885669 0.679104 0.639816
# Saida 3
 0 1

# Entrada  4
0.083714 0.547349 0.451844 0.448608 0.270405 0.507507 0.525222 0.056247 0.521891 0.499813
0.605503 0.507447 0.513353 0.506610 0.482002 0.475198
# Saida 4
 1 0
```

Figura 6.2: Exemplo de um arquivo .pat.

sumo dos experimentos realizados. A coluna **Perfis** indica a combinação de perfis utilizada na entrada da rede, a coluna N^o **de Neurônios na Camada Escondida** indica o número de neurônios da camada escondida, a coluna **SSE/Número de Saídas** indica média do erro de treinamento e a coluna **Taxa de Acertos** representa a porcentagem de acertos no conjunto de treinamento. O erro foi definido como sendo a medida da soma dos quadrados das diferenças entre as saídas reais e desejadas (SSE - *Sum of Squared Errors*), conforme a Equação 6.5. Em todas as redes foram realizados 100.000 passos de treinamento. Os números de neurônios apresentados na Tabela 6.5 foram os que apresentaram melhor resultado para estes conjuntos.

Perfis	N^o **de Neurônios na Camada Escondida**	**SSE/Número de Saídas**	**Taxa de Acertos**
GR, DT	24	3,69090	80 %
DT, ILD	11	7,45037	50,83 %
DT, NPHI	50	3,68075	78,96 %
DT, RHOB	40	2,96798	83,33 %
GR, NPHI	50	3,21602	82,92 %
GR, ILD	28	5,42403	65,21 %
GR, RHOB	40	4,86898	64,37 %
GR, ILD, RHOB	24	2,82767	83,96 %
ILD, NPHI	24	6,63069	67,5 %
ILD, RHOB	32	6,16440	65,62 %

Tabela 6.5: Resumo dos experimentos realizados.

$$SSE = \sum_{i=1}^{n} (SR_i - SD_i)^2 \tag{6.5}$$

em que SR_i e SD_i representam as saídas reais e desejadas de cada padrão i, respectivamente, e n é o número de padrões no conjunto de treinamento.

Nestes conjuntos foram selecionados o mesmo número de padrões em todos os poços, de acordo com o poço que possuia o menor número de padrões. Desta forma, haviam 480 padrões nos conjuntos de treinamento. Para os conjuntos com dois perfis, haviam 7 neurônios na camada de entrada e para o conjunto com 3 perfis haviam 10 neurônios na camada de entrada.

Para o mesmo conjunto de dados foi realizado um experimento contendo somente os perfis e a profundidade como dados de entrada, isto é, sem as inclinações das curvas. Para este caso, o número de neurônios da camada de entrada foi 6 e na camada escondida 15. A taxa de acerto no conjunto de treinamento foi de 72,71 %. Também foi realizado um experimento com apenas uma classe, ou seja, uma classe possuia o valor 1 e as demais 0. Neste caso, a taxa de acerto no conjunto de treinamento foi de apenas 15 %.

Com estes experimentos, observou-se que o erro máximo de treinamento esperado (da ordem de 0.1) nunca foi alcançado e por consequência os experimentos de teste não chegaram a ser realizados.

Além disso, foram realizados experimentos com dados de apenas dois poços do Campo de Namorado (3NA 0001A RJS e 3NA 0004 RJS). Primeiro, foram apresentados à rede os padrões de apenas dois tipos de litofácies (8 - Arenito Médio Maciço Gradado e 16 - Interlaminado Siltito e Folhelho Bioturbado), depois de três tipos (8, 16 e 10 - Arenito / Folhelho Interestratificado) e por fim de quatro tipos (8, 16, 10 e 12 - Siltito Argiloso Estratificado). Estas litofácies foram escolhidas porque, nestes poços, eram as que possuiam maior número de padrões. Os resultados mostraram que quanto maior o número de litofácies piores as taxas de acerto do conjunto de teste, conforme mostra a Tabela 6.6.

Litofácies	**Taxa de Acerto**
(8) vs (16)	90,91%
(8) vs (16) vs (10)	72,09%
(8) vs (16) vs (10) vs (12)	50,94%

Tabela 6.6: Taxas de acerto conforme o acréscimo de litofácies. vs = versus

Após uma consulta a um geólogo ficou constatado que isto ocorria devido ao grande nível de detalhamento dos testemunhos. Para solucionar este problema, um novo pré-processamento foi realizado, que consistiu em reduzir o número de litofácies a partir de uma análise de agrupamentos, uma vez que acreditamos que estas dificuldades no treinamento estão associadas ao número excessivo de litofácies. Estes experimentos serão apresentados na próxima seção.

Os resultados dos experimentos iniciais e dos experimentos com perfis de Potencial Espontâneo foram descritos e discutidos em um artigo e submetido ao Workshop de Teses

e Dissertações em Inteligência Artificial, que é parte do Simpósio Brasileiro de Inteligência Artificial. O artigo foi aceito e será publicado nos anais do evento que acontecerá em novembro de 2002 em Porto de Galinhas - Recife/PE [Cunha and Gomes, 2002].

6.3 Experimentos com Agrupamento de Litofácies

Os experimentos finais da dissertação se caracterizam principalmente pela eliminação de litofácies, pelo agrupamento de litofácies e pelo tratamento de padrões problemáticos. Nesta fase, somente 8 poços fizeram parte dos experimentos, sendo 5 para treinamento e 3 para teste. Isto ocorreu porque um dos poços utilizados na fase anterior, continha apenas uma pequena parte dos dados de testemunho legível. Com a diminuição do número de poços, o número total de litofácies disponíveis para experimentação também diminuiu, pois algumas litofácies só ocorriam no poço eliminado. Desta forma, restaram 22 litofácies. Nos experimentos anteriores foram utilizadas todas as 28 litofácies. As litofácies eliminadas foram: 23, 24, 25, 26, 27 e 28.

Primeiramente, todas as 22 litofácies foram utilizadas. A rede era composta por 16 neurônios na camada de entrada, 16 neurônios na camada escondida e 22 neurônios na camada de saída. O conjunto de dados era composto por 1354 padrões de treinamento e 740 padrões de teste. A Tabela 6.7 mostra os resultados do treinamento e teste da Rede Neural para este conjunto de dados. A taxa de acerto no conjunto de treinamento não foi de 100 % porque mesmo após um grande número de interações a rede não convergia. Observou-se que o erro havia estabilizado, portanto, optou-se por parar o treinamento.

Taxa de Acerto no Conjunto de Treinamento	67,65 %
Taxa de Acerto no Conjunto de Teste	11,08 %

Tabela 6.7: Resultado do treinamento e teste com todas as litofácies.

Este resultado inicial ocorreu devido ao alto nível de detalhamento das litofácies, à imprecisão da associação dos testemunhos aos perfis e ao diferente número de padrões de cada litofácie. Para resolver estes problemas, decidiu-se fazer um agrupamento das litofácies relacionadas. O agrupamento realizado é justificável devido às propriedades das rochas pertencentes ao mesmo grupo serem semelhantes. Estes grupos são conhecidos como eletrofá-

cies, termo discutido na seção 6.3 deste capítulo. Para os experimentos que se seguem a arquitetura da rede tinha 16 neurônios na camada de entrada, 10 neurônios na camada escondida e foram realizados 100000 passos de treinamento. Utilizou-se um número arbitrário de neurônios na camada escondida porque, no momento, não estávamos interessados em descobrir qual a melhor arquitetura da rede. Nosso principal objetivo era investigar se o agrupamento de litofácies seria uma boa estratégia.

Nesta fase, além das litofácies eliminadas por consequência da retirada de um poço, outras foram excluídas por serem consideradas prejudiciais aos experimentos (litofácies 1, 2 e 18). As litofácies são consideradas prejudiciais quando os valores de seus perfis variam muito dentro de um intervalo de profundidade da mesma classe. Estas litofácies foram excluídas por sugestão de um geólogo. Desta forma, das 22 litofácies, restaram 19.

Balanceando os Conjuntos de Treinamento e Teste

Para iniciar os experimentos, as litofácies foram escolhidas de acordo com seus números de ocorrências, ou seja, as primeiras litofácies a fazerem parte dos conjuntos de treinamento e teste foram as que ocorriam com maior frequência (litofácie 8 versus litofácie 17). A Tabela 6.8 mostra o número total de exemplos de cada litofácie nos poços selecionados.

Para o conjunto de dados com apenas duas litofácies haviam 509 padrões de treinamento e 327 padrões de teste. A camada de saída da rede tinha 2 neurônios, cada um representando uma classe. A Tabela 6.9 mostra a média da soma dos erros quadrados (**SSE/Número de Neurônios de Saída**) após o término do treinamento, a taxa de acerto no conjunto de treinamento (**Acerto no Treinamento**) e a taxa de acerto no conjunto de teste (**Acerto no Teste**) em porcentagem.

Padrões Problemáticos

Como estratégia para a identificação de padrões problemáticos, observamos a taxa de acerto para o conjunto de treinamento após um número fixo (tipicamente 100000) de iterações. Caso esta taxa não fosse de 100 %, os padrões não treinados eram identificados, sendo estes candidatos a receberem um tratamento especial.

A taxa de acerto no conjunto de treinamento não foi de 100 % porque a rede não conseguiu treinar um padrão. Ao consultar os dados originais, observou-se que este padrão

Litofácies	**Número de Exemplos**
3	131
4	41
5	11
6	133
7	25
8	491
9	75
10	67
11	102
12	143
13	198
14	34
15	59
16	79
17	345
19	9
20	78
21	70
22	5

Tabela 6.8: Número de exemplos em cada litofácie.

pertencia a uma zona de transição de litofácies. Quando as camadas da formação são muito finas, estas podem sofrer interferência das camadas adjacentes, fazendo com que suas propriedades sejam semelhantes. Esta interferência também ocorre nas zonas de transição de uma camada para outra. Como estavam sendo utilizadas apenas duas litofácies, a litofácie adjacente ao padrão problemático não estava no conjunto. Para resolver este problema o padrão foi removido e a rede foi treinada novamente. Os resultados são mostrados na Tabela 6.10. Percebe-se que após a remoção do padrão problemático, o erro médio no treinamento reduziu drasticamente, assim como as taxas de acerto aumentaram (ver Tabelas 6.9 e 6.10).

Litofácies	(8) vs (17)
SSE/Neurônios de Saída	1,00053
Acerto no Treinamento	99,80 %
Acerto no Teste	67,89 %

Tabela 6.9: Resultado do treinamento com 2 litofácies. vs = versus

Isto indica que uma fase de pré-processamento para remover estes tipos de padrões se faz necessária. Nos próximos experimentos este tratamento sempre se fará presente.

Litofácies	(8) vs (17)
SSE/Neurônios de Saída	0,00107
Acerto no Treinamento	100 %
Acerto no Teste	70,03 %

Tabela 6.10: Resultado do treinamento com 2 litofácies após a remoção do padrão problemático. vs = versus

Agrupamento Incremental de Litofácies

A partir deste ponto discutiremos uma estratégia para agrupar litofácies com base no treinamento de RNAs. Após a rede ter treinado com uma taxa de acerto de 100 % utilizando as litofácies (8) versus (17), a terceira classe mais populosa foi incluída no conjunto de treinamento (litofácie 13). Foram realizados dois treinamentos, um com as três litofácies pertencendo à classes diferentes e outro com a terceira litofácie agrupada com a litofácie 17. A terceira litofácie não foi treinada no mesmo grupo da litofácie 8 devido ao fato de representarem tipos de rochas muito heterogêneas (litofácie 8 - Arenito Médio Maciço Gradado e litofácie 13 - Interlaminado Siltito Argiloso e Marga). Os resultados são mostrados na Tabela 6.11. Este conjunto de dados era composto por 613 padrões de treinamento e por 421 padrões de teste.

Embora a rede não tenha sido treinada com 100 % de acerto no conjunto de treinamento, pode observar-se que a rede obteve melhor resultado quando a terceira litofácie foi agrupada com a litofácie 17. Para o conjunto de treinamento contendo o melhor agrupamento de litofá-

Litofácies	(8) vs (17) vs (13)	(8) vs (17 e 13)
SSE/Neurônios de Saída	4,67121	9,00256
Acerto no Treinamento	97,88 %	98,53 %
Acerto no Teste	52,49 %	78,38 %

Tabela 6.11: Resultado do treinamento com 3 litofácies, antes do tratamento dos padrões problemáticos. vs = versus

cies (conforme Tabela 6.11), constatou-se que alguns dos padrões problemáticos ou estavam em uma zona de transição, ou pertenciam a uma camada muito fina. Consequentemente, decidiu-se alterar a classe associada a esses padrões problemáticos de forma que passassem a pertencer à classe adjacente, ou seja, incluímos um método adicional para tratar os padrões problemáticos (re-classificação), além da pura e simples remoção do padrão. Os padrões problemáticos possuindo classes adjacentes (litofácies) que não pertenciam ao conjunto de treinamento foram removidos. A Tabela 6.12 mostra o resultado do treinamento após o conjunto de treinamento ter sofrido as alterações necessárias.

Litofácies	(8) vs (17 e 13)
SSE/Neurônios de Saída	0,00223
Acerto no Treinamento	100 %
Acerto no Teste	79,10 %

Tabela 6.12: Resultado do treinamento com 3 litofácies agrupadas em 2 classes, após o tratamento dos padrões problemáticos. vs = versus

Após o tratamento de padrões problemáticos pode-se observar, através da comparação entre as Tabelas 6.11 e 6.12, que o erro médio no treinamento diminui drasticamente e as taxas de acerto aumentaram, reforçando a importância do tratamento de padrões problemáticos.

Quando a litofácie 12 (quarta litofácie mais populosa no conjunto de dados) foi inserida no conjunto de treinamento, passaram a existir quatro possibilidades de agrupamento: formar um novo grupo a partir da melhor combinação anterior ((8) versus (17 e 13) versus (12)), agrupar as litofácies 13 e 12 ((8) versus (17) versus (13 e 12)), agrupar as litofácies 17 e

12 ((8) versus (17 e 12) versus (13)) ou agrupar a litofácie 12 ao grupo com as litofácies 17 e 13 ((8) versus (17, 13 e 12)). A possibilidade da litofácie 12 ser agrupada à litofácie 8 não foi considerada devido ao fato de representarem tipos de rochas muito heterogêneas (litofácie 8 - Arenito Médio Maciço Gradado e litofácie 12 - Siltito Argiloso Estratificado). O resultado dos treinamentos realizados são mostrados na Tabela 6.13. Para este conjunto de dados, haviam 749 padrões de treinamento e 428 padrões de teste.

Litofácies	(8) vs (17,12) vs (13)	(8) vs (17) vs (13,12)	(8) vs (17,13) vs (12)	(8) vs (17,13,12)
SSE/Neurônios de Saída	7,95949	12,00534	7,67404	12,00236
Acerto no Treinamento	85,31 %	96,53 %	98,13 %	98,40 %
Acerto no Teste	56,07 %	58,88 %	73,13 %	83,41 %

Tabela 6.13: Resultado do treinamento com 4 litofácies, antes do tratamento dos padrões problemáticos. vs = versus

O melhor resultado de treinamento foi o que continha a litofácie 8 em um grupo e as demais em outro grupo. Após o tratamento dos padrões problemáticos, ou seja, remoção de padrões ou re-classificação, para o melhor conjunto de dados a rede foi treinada novamente e os resultados são apresentados na Tabela 6.14.

Litofácies	(8) vs (17, 13 e 12)
SSE/Neurônios de Saída	0,00204
Acerto no Treinamento	100 %
Acerto no Teste	74,53 %

Tabela 6.14: Resultado do treinamento com 4 litofácies agrupadas em 2 classes, após o tratamento dos padrões problemáticos. vs = versus

Ao comparar as Tabelas 6.13 e 6.14 com relação ao melhor resultado de agrupamento (última coluna da Tabela 6.13 observa-se que, neste caso, embora a taxa de acerto no conjunto de treinamento tenha alcançado 100 % e o erro tenha diminuído drasticamente, a taxa de acerto no conjunto de teste diminuiu consideravelmente após o tratamento dos padrões problemáticos. Provavelmente isto ocorreu devido a um número insuficiente de neurônios

na camada escondida para representar a grande variabilidade das classes. Decidiu-se investigar se o aumento do número de neurônios na camada escondida alteraria as taxas de acerto. A Tabela 6.15 mostra a taxa de acerto no conjunto de teste após o treinamento de uma rede com 16 neurônios na camada escondida (6,78 % a mais do que no último experimento) e após o tratamento dos padrões problemáticos (que foram os mesmos do treinamento anterior).

Litofácies	(8) vs (17, 13 e 12)
SSE/Neurônios de Saída	0,00206
Acerto no Treinamento	100 %
Acerto no Teste	81,31 %

Tabela 6.15: Resultado do treinamento com 4 litofácies agrupadas em 2 classes, após o tratamento dos padrões problemáticos, com 16 neurônios na camada escondida da Rede Neural. vs = versus

Desta forma, ao comparar as Tabelas 6.14 e 6.15, confirma-se a hipótese anteriormente levantada. Ao acrescentar neurônios na camada escondida da Rede Neural o poder de generalização aumentou, fazendo com que a taxa de acerto aumentasse também.

Quando a quinta litofácie foi inserida no conjunto de treinamento (litofácie 6), haviam apenas duas possibilidades de agrupamento (considerando o último melhor agrupamento): formar um novo grupo contendo apenas a nova litofácie ((8) versus (17, 13 e 12) versus (6)) ou agrupar as litofácies 8 e 6 ((8 e 6) versus (17, 13 e 12)). Os resultados destas possibilidades estão na Tabela 6.16. Da mesma forma que nos conjuntos anteriores, não foram realizados experimentos com as outras possibilidades de agrupamento devido aos tipos de rocha muito heterogêneas. Para este grupo, haviam 808 padrões de treinamento e 502 padrões de teste.

Litofácies	(8) vs (17, 13 e 12) vs (6)	(8 e 6) vs (17, 13 e 12)
SSE/Neurônios de Saída	10,01730	7,00352
Acerto no Treinamento	97,77 %	99,13 %
Acerto no Teste	70,92 %	82,07 %

Tabela 6.16: Resultado do treinamento com 5 litofácies, antes do tratamento dos padrões problemáticos. vs = versus

O melhor resultado foi agrupando a litofácie 6 com a litofácie 8 e mantendo as demais em outro grupo. A Tabela 6.17 mostra o resultado do treinamento da rede após o tratamento dos padrões problemáticos.

Litofácies	(8 e 6) vs (17, 13 e 12)
SSE/Neurônios de Saída	0,00337
Acerto no Treinamento	100 %
Acerto no Teste	86,65 %

Tabela 6.17: Resultado do treinamento com 5 litofácies após o tratamento dos padrões problemáticos. vs = versus

Nas Tabelas 6.16 e 6.17 pode-se observar que após o tratamento de padrões problemáticos o erro médio diminuiu drasticamente e as taxas de acerto aumentaram, da mesma forma que para os agrupamentos anteriores.

Após a inserção destas litofácies no conjunto de treinamento, mais duas litofácies foram inseridas, desta vez sem tentar várias combinações de agrupamento. Isto foi feito com base na experiência adquirida durante o agrupamento das outras litofácies. As litofácies 7 e 9 foram agregadas às litofácies 8 e 6. As taxas de acerto no conjunto de treinamento e teste são mostrados na Tabela 6.18.

Litofácies	(8, 6, 7 e 9) vs (17, 13 e 12)
SSE/Neurônios de Saída	11,00139
Acerto no Treinamento	98,73 %
Acerto no Teste	82,47 %

Tabela 6.18: Resultado do treinamento com 7 litofácies, antes do tratamento dos padrões problemáticos. vs = versus

A Tabela 6.19 mostra o resultado do treinamento deste conjunto de litofácies após o tratamento dos padrões problemáticos. O erro máximo de treinamento também é mostrado. O conjunto de treinamento possuia 868 padrões e o conjunto de teste possuia 543 padrões.

Da mesma forma que na maioria dos grupos, o erro médio diminuiu consideravelmente e as taxas de acerto no conjunto de treinamento e teste aumentaram, conforme indicam as

Litofácies	(8, 6, 7 e 9) vs (17, 13 e 12)
SSE/Neurônios de Saída	0,00362
Acerto no Treinamento	100 %
Acerto no Teste	83,95 %

Tabela 6.19: Resultado do treinamento com 7 litofácies após o tratamento dos padrões problemáticos. vs = versus

Tabelas 6.18 e 6.19.

A partir deste conjunto, quando outra litofácie foi inserida (litofácie 3), o resultado não foi satisfatório, uma vez que a taxa de acerto diminuiu para abaixo de 50 %, conforme mostra a Tabela 6.20. Este resultado ocorreu porque o número de padrões nos conjuntos formados ficou muito grande, de forma que a rede não consegue treinar a litofácie 3 devido à pouca quantidade de padrões disponíveis. Desta forma, devido ao desbalanceamento causado pelos poucos dados disponíveis para as litofácies restantes na base de dados, decidiu-se não inserir mais litofácies nos conjuntos.

Litofácies	(8, 6, 7 e 9) vs (17, 13 e 12) vs (3)
SSE/Neurônios de Saída	0,33474
Acerto no Treinamento	99,53 %
Acerto no Teste	48,78 %

Tabela 6.20: Resultado do treinamento com 8 litofácies. vs = versus

Eletrofácies

Os agrupamentos de litofácies realizados durante os experimentos são conhecidos na literatura como eletrofácies [Serra, 1989]. As eletrofácies são um conjunto de respostas de perfis que caracterizam uma camada e permitem que ela seja distinguida das outras. A classificação dos perfis dos poços em eletrofácies não requer qualquer subdivisão artificial da população de dados, ou seja, acontece naturalmente com base nas características únicas das medidas dos perfis do poço que refletem minerais e litofácies dentro de um intervalo perfilado [Lee and Datta-Gupta, 1999].

É possível verificar a coerência dos grupos formados nos experimentos através de uma análise das litofácies agrupadas. A Tabela 5.1 (Subseção 5.3.2 do Capítulo 5) apresenta os nomes das litofácies utilizadas nos experimentos. Nesta tabela, pode-se identificar que foi criada uma eletrofácie composta por arenitos (8, 6, 7 e 9) e outra composta por interestratificados de siltito com marga (17, 13 e 12).

Estes agrupamentos podem ser justificados pelo fato de que alguns tipos de rochas são permeáveis e porosas o suficiente para permitir o acúmulo de petróleo e outras não. A rocha que permite o acúmulo de petróleo é chamada reservatório. Desta forma, os arenitos e calcarenitos podem se constituir rochas-reservatório, além de todas as rochas sedimentares essencialmente dotadas de porosidade intergranular que sejam permeáveis.

Eliminando Padrões Problemáticos dos Conjuntos de Teste

A partir deste ponto, levou-se em consideração apenas o conjunto que obteve maior taxa de acerto no conjunto de teste, ou seja, o grupo com as litofácies (8 e 6) versus (17, 13 e 12), com os padrões problemáticos do conjunto de treinamento tratados. Decidiu-se fazer o mesmo tratamento realizado no conjunto de treinamento para os padrões problemáticos do conjunto de teste. Assim, a taxa de acerto no conjunto de teste sofreu um pequeno aumento. A Tabela 6.21 mostra o resultado deste tratamento.

Litofácies	**Acerto no Teste**
(8 e 6) vs (17, 13 e 12)	89,45 %

Tabela 6.21: Taxa de acerto no conjunto de teste do grupo contendo as litofácies (8 e 6) vs (17, 13 e 12) após o tratamento dos padrões problemáticos no conjunto de teste. vs = versus

Escolhendo a Melhor Arquitetura Neural

Após estes experimentos decidiu-se aumentar o número de neurônios da camada escondida para verificar se a taxa de acerto no conjunto de testes aumentaria. Nestes experimentos as camadas de entrada e escondida eram compostas por 16 neurônios cada e a camada de saída por 2 (nos experimentos anteriores a camada escondida era composta por 10 neurônios). Os agrupamentos foram mantidos e os padrões problemáticos do conjunto de treinamento tratados. A Tabela 6.22 mostra o resultado destes experimentos.

Litofácies	Acerto no Teste
(8 e 6) vs (17, 13 e 12)	85,66 %

Tabela 6.22: Taxa de acerto no conjunto de teste após o aumento do número de neurônios na camada escondida e tratamento dos padrões problemáticos. vs = versus

Ao comparar as Tabelas 6.21 e 6.22 percebe-se portanto que um aumento na complexidade da Rede Neural não proporcionou uma melhora na taxa de classificação e sim uma redução. Isto provavelmente se deve a um fenômeno chamado de super-especialização (ou *overfiting*), ou seja, a rede deixa de generalizar os padrões e passa a representar internamente cada padrão como uma classe, devido ao alto número de neurônios.

O papel das Inclinações das Curvas nos Gráficos dos Perfis

Em seguida, com o objetivo de investigar a importância das inclinações nas curvas dos gráficos dos perfis no processo de reconhecimento, decidiu-se retirar do conjunto de treinamento e teste as inclinações das curvas nos gráficos. A Tabela 6.23 mostra o resultado sem as inclinações das curvas. As Redes Neurais para estes conjuntos eram formadas por 6 neurônios nas camadas de entrada e escondida e 2 neurônios na camada de saída. Da mesma forma que nos experimentos iniciais com o perfil Potencial Espontâneo, decidiu-se que o número de neurônios da camada escondida deveria ser o mesmo da camada de entrada.

Litofácies	Acerto no Treinamento	Acerto no Teste
(8 e 6) vs (17, 13 e 12)	99,25 %	80,68 %

Tabela 6.23: Taxa de acerto no conjunto de teste e treinamento após a remoção das inclinações das curvas. vs = versus

A rede não atingiu 100 % de acerto para o conjunto de treinamento porque não treinaram alguns padrões não problemáticos (ou seja, que não estavam nem em zonas de transições nem em camadas muito finas). Neste caso, os padrões foram mantidos e as taxas de acerto não alcançaram um valor satisfatório. Ao comparar as Tabelas 6.21 e 6.23 observa-se que a taxa de acerto no conjunto de teste diminuiu considervelmente ao retirarmos as inclinações das curvas, o que comprova a importância destas variáveis.

Comparação entre Todas as Técnicas

Todos estes experimentos com os mesmos conjuntos de dados e com técnicas diferentes foram realizados para verificar qual técnica se aplicaria melhor. A título de comparação, a Tabela 6.24 mostra as taxas de acerto nos conjuntos de teste para todas as técnicas utilizadas com os dados do Campo Escola de Namorado.

Litofácies	(8) vs (17)	(8) vs (17 e 13)	(8) vs (17, 13 e 12)	(8 e 6) vs (17, 13 e 12)	(8, 6, 7 e 9) vs (17, 13 e 12)
com inclinações	70,03 %	79,10 %	74,53 %	86,65 %	83,95 %
com mais neurônios	67,89 %	78,15 %	81,31 %	85,66 %	81,55 %
sem inclinações	67,89 %	89,07 %	79,91 %	80,68 %	80,26 %

Tabela 6.24: Taxas de acerto nos conjuntos de teste para todas as técnicas. vs = versus

Para o conjunto com as litofácies 8 e 17, pode-se observar através da Tabela 6.24 que a melhor técnica é utilizar as inclinações das curvas nos gráficos. Para o conjunto com as litofácies 8, 17 e 13, a melhor técnica é retirar as inclinações das curvas dos conjuntos de treinamento e teste. Para o conjunto com as litofácies 8, 17, 13 e 12, a melhor técnica é acrescentar mais neurônios na camada escondida e deixar as inclinações das curvas nos conjuntos de treinamento e teste. Para os dois últimos conjuntos a melhor técnica é simplesmente deixar as inclinações das curvas nos gráficos.

É possível identificar que, para os agrupamentos maiores de litofácies (duas últimas colunas da Tabela 6.24), a técnica que obteve melhores resultados foi a que utilizou as inclinações das curvas dos gráficos dos perfis. Portanto, acredita-se que para conjuntos maiores de eletrofácies, esta técnica seja a mais promissora. Adicionalmente, quando o número de padrões é pequeno, por consequência de um número menor de litofácies, a confiabilidade dos resultados pode ser duvidosa, o que justifica a não uniformidade dos percentuais das várias técnicas para as três primeiras colunas da Tabela 6.24.

Algumas combinações de litofácies plausíveis do ponto de vista geológico podem gerar taxas de reconhecimento elevadas com o uso de RNAs; o método proposto, com os níveis de desempenho e reconhecimento apresentados na Tabela 6.24, pode ser usado na indústria do petróleo como ferramenta de suporte/apoio à decisão. No entanto, na forma em que o conhecimento se encontra disponível (codificado nos pesos da rede), ainda não serve como

ferramenta prática para descoberta de novos conhecimentos, portanto um trabalho futuro será a extração de regras das redes treinadas.

6.4 Sumário

Neste capítulo foi apresentada a parte experimental do trabalho e seus resultados. Algumas diferentes técnicas de treinamento foram utilizadas e analisadas, são elas: treinamento com inclinações das curvas nos gráficos dos perfis acrescidas ao conjunto de treinamento; treinamento padrão, ou seja, sem as inclinações das curvas; treinamento com diferentes arquiteturas da rede; apresentação de uma janela dos dados à rede e agrupamento de classes. No próximo capítulo serão discutidas as conclusões finais do trabalho e perspectivas de trabalhos futuros.

Capítulo 7

Conclusões

Esta dissertação apresentou a proposta de um método para identificação automática de litofácies de poços de petróleo utilizando uma abordagem baseada em Redes Neurais. A identificação de litofácies é uma das etapas fundamentais do processo de caracterização de um reservatório. A caracterização de reservatórios é muito importante no processo de avaliação econômica de um poço. Assim, fica evidente a necessidade de pesquisas relacionadas a este tema.

O método proposto foi implementado e validado conforme os experimentos apresentados e discutidos no capítulo anterior. Nas próximas seções serão apresentados um breve sumário da dissertação, algumas considerações gerais dos resultados obtidos, proposta de trabalhos futuros e algumas considerações finais.

7.1 Sumário da Dissertação

O principal objetivo do primeiro capítulo foi o de apresentar ao leitor, noções do processo de exploração de petróleo e introduzir a técnica de Redes Neurais. Neste capítulo também foram apresentadas algumas áreas de pesquisa de interesse da indústria petrolífera. Vale a pena lembrar que as áreas que apresentamos são apenas algumas dentre muitas outras. Nesta introdução, Redes Neurais foram apresentadas como uma abordagem conexionista da Inteligência Artificial. Além disto, os objetivos e relevância do trabalho foram explicitados.

No Capítulo 2 foram apresentados conceitos fundamentais sobre a formação de um reservatório, os constituintes do petróleo e das etapas de perfuração que produziram os dados

utilizados neste trabalho: perfilagem e testemunhagem. Estes conceitos são importantes para que se possa compreender a necessidade de um processo automático de identificação de litofácies. Por possuir um custo elevado, a testemunhagem não é realizada para todos os poços, embora seja muito mais precisa na identificação de litofácies que a perfilagem. Visto que a perfilagem está disponível para todos os poços é interessante realizar a identificação de litofácies a partir destes dados. Como a análise manual de perfis consome muito tempo e requer um conhecimento específico, é interessante ter um processo automático de identificação de litofácies. Com o problema definido e com um embasamento sobre as variáveis que o compõem, partiu-se para a conceituação básica da técnica utilizada, Redes Neurais em descoberta de conhecimento.

O Capítulo 4 apresentou uma revisão bibliográfica englobando os principais tópicos do trabalho. Neste capítulo, foram apresentadas técnicas convencionais para identificação de litofácies (abordagens estatísticas) e trabalhos utilizando Redes Neurais para esta aplicação. Este trabalho foi inspirado em um dos artigos apresentados [White et al., 1995]. Também foram apresentadas as principais técnicas de extração de regras de Redes Neurais, cuja análise nos levou a sugerir a aplicação do algoritmo FERNN (Apêndice A) à saída da rede. As várias aplicações de Redes Neurais na indústria do petróleo foram apresentadas ao final do capítulo.

O Capítulo 5 enfocou a importância da identificação automática de litofácies, bem como propôs um método para esta aplicação. O método proposto é composto pelas etapas de associação dos dados de perfis aos dados de testemunho, discretização dos dados, agrupamento de classes, treinamento da Rede Neural, tratamento de padrões problemáticos, extração de regras e validação. As três primeiras etapas compõem o pré-processamento. Os dados do campo Escola de Namorado foram apresentados neste capítulo, dando destaque aos dados selecionados para a fase de experimentação.

O Capítulo 6 apresentou detalhadamente os experimentos e resultados neste trabalho. Neste capítulo, procuramos mostrar em ordem cronológica a evolução dos experimentos. Na próxima seção serão feitas algumas considerações sobre a dissertação.

7.2 Considerações Gerais

Redes Neurais Artificiais são uma técnica pouco utilizada em descoberta de conhecimento devido a três fatores principais:

- As Redes Neurais necessitam de vários passos sobre o conjunto de dados para obter uma alta taxa de classificação, consumindo muito tempo no treinamento.
- Como as informações são representadas por pesos de múltiplas conexões, articular regras compreensíveis se tornam um grande problema.
- Pelo mesmo motivo anterior, é muito difícil incluir novos conhecimentos à rede.

No entanto, o principal motivo que nos levou a estudar esta técnica é que as Redes Neurais possuem melhor desempenho quando está tratando dados com ruído do que outras técnicas. Além disso, RNAs possuem alta capacidade de generalização e aprendizagem a partir de exemplos.

Existe uma grande diversidade de aplicações de Redes Neurais à indústria petrolífera, dentre elas o problema de identificação de litofácies de reservatórios de petróleo. Um dos principais motivos que nos levou a escolher este problema é que os trabalhos realizados até o momento utilizavam poucos dados na fase de experimentação (tipicamente 5 poços). Além disso, não encontramos nenhum trabalho semelhante utilizando dados nacionais, o que torna o trabalho ainda mais relevante.

A escolha do algoritmo (FERNN) de extração de regras ocorreu principalmente devido à sua simplicidade em relação aos outros algoritmos estudados. Na maioria dos casos, os algoritmos requerem um re-treinamento da Rede Neural, o que leva a um consumo ainda maior na etapa de aprendizagem.

As várias técnicas de treinamento discutidas no capítulo de experimentos mostraram que há uma grande dificuldade em automatizar o processo de identificação de litofácies. Isto ocorre principalmente porque há uma incerteza na associação dos dados e um grande nível de detalhamento dos mesmos. As taxas finais de reconhecimento foram consideradas satisfatórias devido à complexidade dos dados e mostraram que Redes Neurais são técnicas promissoras. No entanto, para um melhor desempenho do método outros trabalhos devem ser realizados, os quais são apresentados na próxima seção.

Durante o trabalho, todas as etapas do método proposto, com exceção da extração de regras, foram implementadas e validadas. A associação dos dados é frequentemente utilizada no processo de identificação de litofácies. A discretização realizada nos dados é comum em aplicações que utilizam Redes Neurais. No entanto, o agrupamento de litofácies em eletrofácies via técnica de treinamento é nova, pois geralmente os agrupamentos são feitos por *Box-plot*. O tratamento de padrões problemáticos também é novidade. A etapa de extração de regras pode ser considerada nova levando em consideração que nunca foi aplicada na identificação de litofácies. Desta forma, o desenvolvimento de tal método foi uma das contribuições importantes da dissertação.

7.3 Trabalhos Futuros

O escopo do trabalho, definido inicialmente como propor um método de identificação automática de litofácies e estudar um algoritmo de extração de regras de Redes Neurais foi concluído com sucesso, pois as taxas de reconhecimento da rede foram consideradas satisfatórias. A etapa de associação dos dados foi realizada manualmente. A etapa de discretização, por questões de simplicidade, foi realizada em uma planilha excel. Para melhorar o método, poderia ser construído um programa de normalização de dados, em que o usuário possa escolher a forma de apresentação de padrões e de normalização. Por exemplo, o usuário poderia indicar qual a função e o intervalo de normalização dos dados e o arquivo de entrada da rede seria criado automaticamente de acordo com as especificações informadas. Além disso, algumas sugestões de trabalhos futuros para seguir a linha de pesquisa são:

- Realizar, com auxílio de um especialista, um tratamento mais preciso nos dados de perfis e testemunhos de forma que não hajam padrões problemáticos desde a primeira vez que a rede treina.
- Extrapolar o número de padrões disponíveis para as litofácies com poucos exemplos utilizando, por exemplo, combinação linear.
- Utilizar Análise dos Componentes Principais (PCA) ou outro método estatístico para agrupar as litofácies.
- Implementar e validar a etapa de extração de regras.

- Investigar a combinação de dados sísmicos, testemunhos e perfis para resolução do problema.

7.4 Considerações Finais

Conforme discutido nesta dissertação, o trabalho desenvolvido engloba duas áreas bem distintas: descoberta de conhecimento e engenharia do petróleo. Durante o processo de formação, várias disciplinas relacionadas à primeira área foram cursadas e vários cursos relacionados à exploração de petróleo foram realizados. As disciplinas cursadas foram: Inteligência Artificial, Mineração de Dados, Banco de Dados, Teoria da Computação, Documentação Técnico-Científica, Sistemas de Informações Geo-referenciadas e Metodologia do Ensino Superior. Os cursos extra-curriculares realizados foram:

- Geologia do Petróleo;
- Perfuração de Poços de Petróleo;
- Noções de Engenharia de Reservatórios e
- Reservatórios de Petróleo / Gás e o Meio Ambiente.

Além das disciplinas e cursos realizados, foi apresentado um seminário no I Encontro do Programa de Petróleo e Gás da ANP/CCT/UFPB [Cunha and Barros, 2000], realizado pelo PRH 25, sobre trabalho colaborativo para apoio à tomada de decisões com bases em sistemas de informações geográficas aplicados à indústria do petróleo. O problema de identificação de litofácies que se propôs investigar utilizando uma abordagem de Redes Neurais, foi apresentado a dois geólogos que confirmaram a importância do trabalho e incentivaram a busca por novas técnicas nesta linha de pesquisa. O resumo desta dissertação foi apresentado no Congresso Nacional de Pós-graduandos realizado em Campina Grande [Cunha and Gomes, 2001].

Bibliografia

[Agrawal and Srikant, 1994] Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules. In Bocca, J. B., Jarke, M., and Zanido, C., editors, *Proc. 20th Int. Conf. Very Large Data Bases (VLDB)*, pages 487–499. Morgan Kaufmann.

[Al-Qahtani, 2000] Al-Qahtani, F. A. (2000). Porosity distribution prediction using artificial neural networks. Master's thesis, College of Engineering and Mineral Resources at West Virginia University.

[Alsabti et al., 1998] Alsabti, K., Ranka, S., and Singh, V. (1998). An efficient k-means clustering algorithm. In *(IPPS): 11th International Parallel Processing Symposium*. IEEE Computer Society Press.

[Aminian et al., 2000] Aminian, K., Bilgesu, H. I., Riera, A., Chamorro, A., and Ameri, S. (2000). Enhancing secondary oil recovery performance simulation with the aid of a neural network. In *IASTED International Conference*. West Virginia University, Pittsburgh, USA.

[Aurélio et al., 1999] Aurélio, M., Vellasco, M., and Lopes, C. H. (1999). Descoberta de conhecimento e mineração de dados. Apostila.ICA - Laboratório de Inteligência Computacional Aplicada, DEE, PUC-Rio.

[Beale and Jackson, 1990] Beale, R. and Jackson, T. (1990). *Neural Computing: An Introduction*. Institute of Physics Publishing Bristol and Philadelphia.

[Bloch et al., 2001] Bloch, M., Couto, M. J., Souza, A. A., and Perez, R. C. (2001). Métodos para a estimativa da duração da perfuração e completação de poços offshore. In *Anais do I Congresso Brasileiro de P & D em Petróleo e Gás*, page 199. Resumo - Natal/RN - Brasil.

[Cendrowska, 1987] Cendrowska, J. (1987). Prism: An algorithm for inducing modular rules. *International Journal of Man-Machine Studies*, 27:349–370.

[Coll et al., 1999] Coll, C., Jing, X. D., and Muggeridge, A. H. (1999). Integration of core and log information to improve the representation of small/medium-scale heterogeneity. *SPE Annual Technical Conference and Exhibition*. SPE 56804.

[Cunha and Barros, 2000] Cunha, E. S. and Barros, M. A. (2000). Trabalho colaborativo para apoio à tomada de decisões com base em sistemas de informação geográfica (sig). Apresentado no I Encontro do Programa de Petróleo e Gás da ANP/CCT/UFPB.

[Cunha and Gomes, 2001] Cunha, E. S. and Gomes, H. M. (2001). Mineração de dados utilizando redes neurais com aplicação ao setor de petróleo e gás. XVI Congresso Nacional de Pós-graduandos e II Encontro dos Pós-graduandos da UFPB.

[Cunha and Gomes, 2002] Cunha, E. S. and Gomes, H. M. (2002). Identificação de litofácies de poços de petróleo utilizando um método baseado em redes neurais. Aceito para publicação no I Workshop de Teses e Dissertações em Inteligêcia Artificial.

[de Carvalho, 2001] de Carvalho, L. A. V. (2001). *Datamining - A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração*. Érica.

[Doveton, 1994] Doveton, J. H. (1994). Geologic log interpretation. SEPM Short Course, No 29. Society for Sedimentary Geology. Tulsa, Oklahoma.

[Duda et al., 2000] Duda, R. O., Hart, P. E., and Stork, D. (2000). *Pattern Classification*. Wiley-Interscience. 2nd edition.

[Fayyad et al., 1996] Fayyad, U. M., Shapiro, G. P., Smyth, P., and Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*. AAAI Press, The MIT Press.

[Fidelis et al., 2000] Fidelis, M. V., Lopes, H. S., and Freitas, A. A. (2000). Discovering comprehensible classification rules with a genetic algorithm. In *Proc. Congress on Evolutionary Computation - 2000 (CEC-2000)*, pages 805–810. La Jolla, CA, USA.

[Filho, 1994] Filho, J. D. S. (1994). Utilização de simulador numérico na análise do processo de migração secundária de petróleo. Master's thesis, Universidade Estadual de Campinas.

[Haykin, 1999] Haykin, S. (1999). *Neural Networks. A Comprehensive Foundation*. Prentice Hall. 2nd edition.

[He et al., 2001] He, Z., Yang, L., Yen, J., and Wu, C. (2001). Neural-network approach to predict well performance using available field data. In *Society of Petroleum Engineers - SPE 68801*.

[Holsheimer and Siebes, 1994] Holsheimer, M. and Siebes, A. (1994). Data mining. the search for knowledge in databases. Technical Report CS-R9406, CWI.

[Hruschka and Ebecken, 1999] Hruschka, E. R. and Ebecken, N. F. F. (1999). Extração de regras de redes neurais por meio do algoritmo rx modificado: Um exemplo de aplicação em modelagem de dados meteorológicos. In *Proceedings of the IV Brazilian Conference on Neural Network - IV Congresso Brasileiro de Redes Neurais. ITA - SP, São José dos Campos*, pages 47–51.

[Hruschka and Ebecken, 2000] Hruschka, E. R. and Ebecken, N. F. F. (2000). Applying a clustering genetic algorithm for extracting rules from a supervised neural network. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN'00)*. Corno, Italy.

[Huang et al., 1996] Huang, Z., Shimeld, J., Williamson, M., and Katsube, J. (1996). Permeability prediction with artificial neural network modeling in the venture gas field, offshore eastern canada. *Geophysics*, 61(2):422–436.

[Jolliffe, 1986] Jolliffe, I. (1986). *Principal Component Analysis*. Springer Verlog. New York.

[Junior and Yoneyama, 2000] Junior, C. L. N. and Yoneyama, T. (2000). *Inteligência Artificial em Controle e Automação*. Edgard Blücher Ltda. 1^a edition.

[Lee and Datta-Gupta, 1999] Lee, S. H. and Datta-Gupta, A. (1999). Electrofacies characterization and permeability predictions in carbonate reservoirs: Role of multivariate analysis and nonparametric regression. *SPE Annual Technical Conference and Exhibition*. SPE 56658.

[Lu et al., 1995] Lu, H., Setiono, R., and Liu, H. (1995). Neurorule: A connectionist approach to data mining. In *Proc.* 21^{st} *Very Large Databases Conf. (VLDB - 95)*, pages 478–489.

[Lu et al., 1996] Lu, H., Setiono, R., and Liu, H. (1996). Effective data mining using neural networks. *Proc. IEEE Transactions on Knowledge and Data Engineering*, 8(6):957–961.

[Luke, 2002] Luke, B. T. (2002). K-means clustering. Disponível on-line em: http://fconyx.ncifcrf.gov/l̃ukeb/kmeans.html.

[Macêdo, 2001] Macêdo, A. R. M. (2001). Sistema aditivo de prevenção contra incidentes com incêndio e danos ao meio ambiente em Área industrial do setor de petróleo. In *Anais do I Congresso Brasileiro de P & D em Petróleo e Gás*, page 248. Resumo - Natal/RN - Brasil.

[McVey et al., 1994] McVey, D., Mohaghegh, S., Aminian, K., and Ameri, S. (1994). Identification of parameters influencing the response of gas storage wells to hydraulic fracturing with the aid of a neural network. In *Proceedings of 1994 SPE Eastern Regional Conference and Exhibition - SPE 29159*, pages 31–39. Charleston, West Virginia.

[Miranda and Baptista, 2001] Miranda, R. A. V. and Baptista, C. S. (2001). Ecolib: Uma biblioteca digital multimídia para gestão de meio ambiente na Área de petróleo e gás. In *Anais do I Congresso Brasileiro de P & D em Petróleo e Gás*, page 237. Resumo - Natal/RN - Brasil.

[Mitchell, 1998] Mitchell, M. (1998). *An Introduction to Genetic Algorithms*. MIT Press.

[Mohaghegh and Ameri, 1995] Mohaghegh, S. and Ameri, S. (1995). Artificial neural network as a valuable tool for petroleum engineers. In *Society of Petroleum Engineers - SPE 29220*.

[Mohaghegh et al., 1995] Mohaghegh, S., Arefi, R., Ameri, S., and Rose, D. (1995). Design and development of an artificial neural network for estimation of formation permeability. In *Proceedings SPE Petroleum Computer Conference - SPE 28237*, pages 151–154. Dallas, Texas.

[Mohaghegh et al., 1996] Mohaghegh, S., McVey, D., Aminian, K., and Ameri, S. (1996). Predicting well stimulation results in a gas storage field in the absence of reservoir data, using neural networks. *SPE Reservoir Engineering Journal - SPE 31159*, pages 54–57.

[Mohaghegh et al., 1998] Mohaghegh, S., Richardson, M., and Ameri, S. (1998). Virtual magnetic imaging logs: generation of synthetic mri logs from conventional well logs. In *Proc. SPE East Reg. Conf. - SPE 51075*, pages 223–232. Pittsburgh.

[Mohn et al., 1987] Mohn, E., Berteig, V., and Helgeland, J. (1987). A review of statistical approaches to lithofacies determination from well data. *North Sea and Gas Reservoirs. The Norwegian Institute of Technology*, pages 301–309.

[Nievola et al., 1999] Nievola, J. C., Santos, R. T., Freitas, A. A., and Lopes, H. S. (1999). Extração de regras de redes neurais via algoritmos genéticos. In *Proceedings of the IV Brazilian Conference on Neural Networks - IV Congresso Brasileiro de Redes Neurais*, pages 158–163. Resumo.

[OWG, 2000] OWG (2000). Data mining. OWG - Smart Business. Smart Solutions. Disponível on-line em: http://www.dwbrasil.com.br/html/dmining.html.

[Pinto and Monteiro, 1998] Pinto, I. G. and Monteiro, V. S. (1998). introdução aos algoritmos genéticos. Universidade Técnica de Lisboa. Disponível on-line em: http://laseeb.ist.utl.pt/portas_abertas/ags/Apend_dncc.html.

[Popp, 1988] Popp, J. H. (1988). *Geologia Geral*. Livros Técnicos e Científicos Editora.

[Quinlan, 1993] Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

[Rich and Knight, 1991] Rich, E. and Knight, K. (1991). *Artificial Intelligence*. McGraw-Hill.

[Russell and Norvig, 1995] Russell, S. and Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall, Englewood Cliffs, NJ.

[Santos et al., 2000] Santos, R. T., Nievola, J. C., and Freitas, A. A. (2000). Extracting comprehensible rules from neural networks via genetic algorithms. In *Proc. 2000 IEEE Symp. on Combinations of Evolutionary Computation and Neural Networks (ECNN-2000)*, pages 130–139. San Antonio, TX, USA.

[SEG, 2002] SEG (2002). Segy format. IRIS PASSCAL Instrument Center. Disponível on-line em: http://www.passcal.nmt.edu/NMT_pages/Software/
segy.shtml.

[Segy, 2002] Segy (2002). Segy format. Disponível on-line em: http://utam.geophys.utah.edu/ebooks/gg522/nmodemo/segy.html.

[Serra, 1989] Serra, O. (1989). *Sedimentary Environments from Wireline Logs*, chapter 5, pages 85–117. Schlumberger.

[Setiono, 2000] Setiono, R. (2000). Extracting m-of-n rules from trained neural networks. *IEEE Transactions on Neural Networks*, 11(2):512–519.

[Setiono and Leow, 1999] Setiono, R. and Leow, W. K. (1999). Generating rules from trained network using fast pruning. In *Proceedings of International Joint Conference on Neural Networks (IJCNN'99)*, pages 4095–4098. Washington D. C.

[Setiono and Leow, 2000] Setiono, R. and Leow, W. K. (2000). Fernn: An algorithm for fast extraction of rules from neural networks. *Applied Intelligence*, 12:15–25.

[Setiono and Liu, 1995] Setiono, R. and Liu, H. (1995). Understanding neural networks via rule extraction. In *Proceedings of the* 14^{th} *International Joint Conference on Artificial Intelligence*, pages 480–485. Montreal, Quebec, Canada: Morgan Kaufmann.

[Silva et al., 2001] Silva, C. T., Neto, M. A. S., Dantas, T. N. C., and Neto, A. A. D. (2001). Estudo das propriedades de um novo fluido de perfuração à base Éster. In *Anais do I Congresso Brasileiro de P & D em Petróleo e Gás*, page 202. Resumo - Natal/RN - Brasil.

[Siripitayananon et al., 2001] Siripitayananon, P., Chen, H., and Hart, B. S. (2001). A new technique for lithofacies prediction: Back-propagation neural network. In *Proceedings of the* 39^{th} *Annual ACM Southeast Conference*. Athens, Georgia.

[SNNS,] SNNS. *Stuttgart Neural Network Simulator*. University of Stuttgart. User Manual, Version 4.2.

[Souza et al., 2001] Souza, C. M. P., Dehárbe, D. P. B., Goldbarg, M. C., and Gouvêa, E. F. (2001). O problema da otimização da locação de poços offshore: Modelagem e solução via transgenética computacional. In *Anais do I Congresso Brasileiro de P & D em Petróleo e Gás*, page 30. Resumo - Natal/RN - Brasil.

[Spinelli et al., 2001] Spinelli, L. S., Junior, D. L. P. M., Louvisse, A. M. T., and Lucas, E. F. (2001). Caracterização das propriedades físico-químicas dos sistemas petróleo/Água: Previsão do comportamento interfacial. In *Anais do I Congresso Brasileiro de P & D em Petróleo e Gás*, page 285. Resumo - Natal/RN - Brasil.

[Tafner et al., 1995] Tafner, M. A., de Xerez, M., and Filho, I. W. R. (1995). *Redes Neurais Artificiais: Introdução e Princípios de Neurocomputação*. Blumenau: EKO: Editora da FURB.

[Thomas, 2001] Thomas, J. E., editor (2001). *Fundamentos de Engenharia de Petróleo*. Interciência.

[Towell and Shavlik, 1993] Towell, G. G. and Shavlik, J. W. (1993). Extracting refined rules from knowledge-based neural networks. *Machine Learning*, 13:71–101.

[White et al., 1995] White, A. C., Molnar, D., Aminian, K., Mohaghegh, S., Ameri, S., and Esposito, P. (1995). The application of ann for zone identification in a complex reservoir. In *Eastern Region Conference Proceedings: Society of Petroleum Engineers*, pages 27–32. SPE 30977.

[Witten and Frank, 1999] Witten, I. H. and Frank, E. (1999). *Data Mining: practical machine learning tools and techniques with Java implementations*. Morgan Kaufmann.

[Yang et al., 2000] Yang, L., Zhong, H., Yen, J., and Wu, C. (2000). A neural network approach to predict existing and infill oil well performance. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN'00)*. Italy.

Apêndice A

Algoritmo de Extração de Regras de Redes Neurais

Este apêndice apresenta o algoritmo para extração de regras de Redes Neurais proposto por Setiono e Leow [Setiono and Leow, 2000; Setiono and Leow, 1999]. O algoritmo consiste dos seguintes passos:

1. Treinar a rede totalmente conectada, tal que uma função de erro *cross-entropy* seja minimizada. A função de erro *cross-entropy* usual é aumentada por uma função de penalização de forma que as conexões irrelevantes e relevantes possam ser diferenciadas pelos seus pesos quando o treinamento termina;

2. Usar os ganhos de informação para identificar as unidades escondidas relevantes e construir uma árvore de decisão que classifique os padrões em termos dos valores de ativação da rede;

3. Para cada unidade escondida cujos valores de ativação são usados como nodo "pai" na árvore de decisão, remover as conexões irrelevantes da entrada para esta unidade escondida. As conexões são irrelevantes se sua remoção não afeta o desempenho de classificação da árvore de decisão;

4. Para conjuntos de dados com atributos discretos, substituir cada nodo "pai" por um conjunto equivalente de regras simbólicas.

O algoritmo de extração de regras de Redes Neurais exige alguns cuidados no momento do treinamento da rede, pois a minimização da função de erro *cross-entropy* garante que as conexões de entradas irrelevantes tenham pesos muito pequenos. Estas conexões podem ser removidas sem afetar a precisão de classificação da rede.

A função de erro *cross-entropy* aumentada é dada por:

$$\theta(w,v) = F(w,v) - \sum_{i=1}^{C}\sum_{p=1}^{P}[t_{ip} log S_{ip} + (1-t_{ip}) log(1-S_{ip})]. \quad \text{(A.1)}$$

$F(w,v)$ é um termo de penalidade

$$F(w,v) = e_1 \sum_{j=1}^{J}\left(\sum_{i=1}^{C}\frac{\beta v_{ij}^2}{1+\beta v_{ij}^2} + \sum_{k=1}^{K}\frac{\beta w_{jk}^2}{1+\beta w_{jk}^2}\right) + e_2 \sum_{j=1}^{J}\left(\sum_{i=1}^{C} v_{ij}^2 + \sum_{k=1}^{K} w_{jk}^2\right) \quad \text{(A.2)}$$

onde e_1, e_2 e β são parâmetros positivos. A função de erro *cross-entropy* tem mostrado ser melhor na convergência do treinamento da rede que as funções de erro *least-squares* padrão. A função de penalidade é adicionada para encorajar o decréscimo dos pesos [Setiono and Leow, 2000]. Através da minimização da função de erro aumentada, espera-se que as conexões inúteis para a classificação dos padrões tenham pesos pequenos e que estas possam ser removidas sem diminuir a qualidade de classificação.

Identificação das Unidades Relevantes

Depois da rede ter sido treinada, suas unidades relevantes são identificadas através do método de ganho de informação. Para isto, o algoritmo C4.5 é aplicado [Quinlan, 1993]. O C4.5 é um algoritmo de classificação baseado em árvores de decisão. Seja T um conjunto de dados e as classes $\{C_1, C_2, ..., C_k\}$, a árvore de decisão é gerada recursivamente pelo C4.5 através dos seguintes passos:

1. Se T contém um ou mais exemplos, todos pertencentes a uma única classe C_j, então a árvore de decisão para T é uma "folha" identificando a classe C_j;

2. Se T é vazio, então a árvore de decisão é uma "folha", onde a classe mais frequente no "pai" deste nodo é escolhida como a classe;

3. Se T contém exemplos que pertencem a várias classes, então o ganho de informação é usado como uma heurística para dividir T em partições baseadas nos valores de uma única característica.

A árvore de decisão é construída usando os valores de ativação dos padrões de treinamento que foram classificados corretamente pela rede, portanto, o número de padrões em T, poderá ser menor que o número de padrões P do conjunto de treinamento. Os valores de ativação são contínuos no intervalo $[0, 1]$ devido ao fato de ter sido usada a função sigmóide dos pesos de entrada para computá-los.

Seja $\overline{P}$ o número de padrões corretamente classificados. Para a unidade escondida j, seus valores de ativação H_{jp} em resposta ao padrão p, $p = 1, 2, ..., \overline{P}$, são ordenados de forma crescente. Os valores de ativação são divididos em dois 2 grupos $T_1 == \{H_{j1,...,H_{jq}}\}$ e $T_2 == \{H_{jq+1}, ..., H_{j,\overline{P}}\}$, onde $H_{jq} < H_{jq+1}$ e o ganho de informação com a divisão é calculado. Este ganho de informação é calculado para todas as possíveis divisões dos valores de ativação, isto é, para q variando de 1 até $\overline{P}$. O ganho máximo será o ganho de informação da unidade escondida j.

Suponha que cada padrão no conjunto de dados T pertença a umas das C classes e n_c seja o número de padrões na classe C_c. A informação esperada para classificação é:

$$Info(T) = -\sum_{c=1}^{C} \frac{n_c}{N} log_2 \frac{n_c}{N} \tag{A.3}$$

onde o número de padrões no conjunto T é $N = \sum_{c=1}^{C} n_c$. Quando aplicado a um conjunto de treinamento, Info(T) mede a quantidade média de informação necessária para identificar a classe de um caso de T (esta quantidade também é conhecida como entropia do conjunto T). Para os dois subconjuntos de T, a informação esperada é computada semelhantemente:

$$I(T_1) = -\sum_{c=1}^{C} \frac{n_{c1}}{N_1} log_2 \frac{n_{c1}}{N_1} \tag{A.4}$$

$$I(T_2) = -\sum_{c=1}^{C} \frac{n_{c2}}{N_2} log_2 \frac{n_{c2}}{N_2} \tag{A.5}$$

onde n_{cj} é o número de exemplos em T_j, $j = 1, 2$ que pertencem a classe C_c e $N_j = \sum_{i=1}^{C} n_{cj}$. O ganho de informação pela divisão de T em T_1 e T_2 é

$$Gain(H_{jq}) = Info(T) - [Info(T_1 + Info(T_2] \tag{A.6}$$

e o ganho normalizado é

$$NGain(H_{jq}) = Gain(H_{jq}/[-\sum_{j=1}^{2}(N_j/N)log_2(N_j/N)] \tag{A.7}$$

O nodo raiz da árvore de decisão contém uma condição de teste que envolve a unidade escondida cujo valor tem o maior ganho normalizado. A árvore de decisão completa é gerada aplicando o mesmo procedimento para os subconjuntos de dados dos dois galhos de um nodo de decisão. Depois que a árvore de decisão foi construída, a identificação das unidades escondidas relevantes é trivial. As unidades escondidas cujas ativações são usadas em um ou mais nodos da árvore de decisão são as unidades relevantes.

Identificação das Conexões de Entrada Relevantes

Devido à rede ter sido treinada com a minimização de uma função de erro que foi aumentada por um termo de penalidade, espera-se que as conexões das entradas irrelevantes tenham pesos pequenos. Para cada unidade escondida j, um ou mais dos pesos de suas conexões w_{jk} das unidades de entrada devem suficientemente pequenos de forma que possam ser removidos sem afetar a precisão de classificação global. O critério para remoção destas conexões irrelevantes é pelas proposições 1 e 2.

Proposição 1 - Seja a condição de teste para um nodo da árvore de decisão igual a $H_{jp} <= H_{jt}$ para algum t. Defina $L == H_{jt}$, $U == H_{j,t+1}$ (o menor valor de ativação que é maior que L), $D_L == \{X_p | H_{jp} = \sigma(W_j X_p) <= L\}$ e $D_U == \{X_p | H_{jp} = \sigma(W_j X_p) >= U\}$. Seja S o conjunto de unidades de entrada cujas conexões para a unidade escondida j satisfazem a seguinte condição:

$$\sum_{k \in S} |w_{jk}| < 2(U - L) \tag{A.8}$$

e seja S' o complemento de S. Então, mudando a condição de teste para

$$H_{jp} <= (L + U)/2, \tag{A.9}$$

as conexões das unidades em S para a unidade escondida j podem ser removidas sem mudar os componentes de D_L e D_U.

Proposição 2 - Seja a condição de teste para um nodo da árvore de decisão igual a $H_{jp} > H_{jt}$ para algum t. Defina $L == H_{jt}$, $U == H_{j,t+1}$ (o menor valor de ativação que é maior que L), $D_L == \{X_p | H_{jp} = \sigma(W_j X_p) <= L\}$ e $D_U == \{X_p | H_{jp} = \sigma(W_j X_p) >= U\}$. Seja S o conjunto de unidades de entrada cujas conexões para a unidade escondida j satisfazem a seguinte condição:

$$\sum_{k \in S} |w_{jk}| <= 2(U - L) \tag{A.10}$$

e seja S' o complemento de S. Então, mudando a condição de teste para

$$H_{jp} > (L + U)/2, \tag{A.11}$$

as conexões das unidades em S para a unidade escondida j podem ser removidas sem mudar os componentes de D_L e D_U.

Após a remoção das conexões redundantes, os nodos que são condição de teste na árvore de decisão devem ser escritos em forma de regras em termos dos pesos da rede.