



2º CONGRESSO BRASILEIRO DE P&D EM PETRÓLEO & GÁS

IDENTIFICAÇÃO AUTOMÁTICA DE LITOFÁCIES EM POÇOS DE PETRÓLEO UTILIZANDO TÉCNICAS AVANÇADAS DE MINERAÇÃO DE DADOS

Josué Toebe, Marcus Costa Sampaio

Universidade Federal de Campina Grande, Rua Otaviano Bezerra da Cunha, 131,
CEP 58108-665, Campina Grande, Paraíba, {toebe,sampaio}@dsc.ufcg.edu.br

Resumo – Um processo de mineração de dados é uma série de passos que objetivam obter conhecimento dos dados. Esses passos incluem a preparação dos dados, seleção de amostras, indução de conhecimento das amostras usando algoritmos de indução e a estimativa da qualidade do conhecimento induzido. A complexidade do processo depende destas várias variáveis relacionadas. Nesta dissertação, nós propomos um processo de mineração e dados automatizado para guiar o minerador em sua tarefa. O processo automatizado considera a existência de uma diversidade de técnicas para cada etapa do processo. Com o objetivo de reduzir os custos, o processo é guiado por heurísticas que ajudam a descobrir a combinação de técnicas que obtém os melhores resultados para um dado banco de dados. Para ilustrar sua aplicação, experimentos que seguem o processo são descritos, utilizando dados do Campo Escola de Namorado, fornecidos pela Agência Nacional do Petróleo do Governo Brasileiro.

Palavras-Chave: Mineração de Dados, Litofácies, Poços de Óleo e Gás Natural.

Abstract – A data mining process is a series of steps to obtain knowledge from data. These steps include random data sampling, sample preparation, knowledge induction from samples using induction algorithms, and estimation of the exactness of the knowledge induced. The complexity of the process is due to these various related variables. In this paper, we propose a process-driven algorithm based on heuristics, to guide the miner through the mining task. It takes into account the diversity of sample selection and preparation techniques, as well as the existence of various induction techniques. The heuristics help to discover the combination of techniques that fits best for a given dataset for mining. To illustrate its application, experiments that follow the process are described, utilizing data from Campo Escola de Namorado, supplied by the Brazilian Government's Petroleum National Agency.

Keywords: Data Mining, Litofácies, Oil and Gas Well.

1. Introdução

O desenvolvimento de técnicas de perfuração de poços para extração de óleo e gás natural propiciou a coleta de grandes quantidades de dados operacionais pelas empresas do ramo. Como o volume de dados tende a crescer exponencialmente com o tempo, aumenta a dificuldade de processar os dados e interpretá-los, visando a tomada de decisões.

Considere a análise de custo/benefício de um reservatório de óleo e gás natural, tendo em vista a sua exploração comercial. Para que a análise seja confiável, é necessário fazer uma descrição detalhada da estrutura litológica do reservatório. A identificação manual de litofácies em um poço de petróleo é um processo intensivo que envolve o dispêndio de uma quantidade considerável de tempo, por parte de um especialista humano experiente. Normalmente, o especialista utiliza, além de seu conhecimento, duas técnicas diferentes para a identificação.

A primeira técnica é denominada *testemunho* e consiste da retirada de amostras de rochas durante o processo de perfuração do poço. De posse da amostra, o especialista faz uma *descrição* do testemunho, identificando de modo preciso as litofácies presentes naquela amostra. A principal desvantagem da utilização desta técnica está no elevado tempo gasto para a retirada da amostra, que em muitos casos obriga a broca de perfuração a ser parada¹, acarretando em um aumento do custo da perfuração.

A segunda técnica é denominada *perfilagem* e consiste da leitura de diversas propriedades físicas das litofácies, com a utilização de sensores. Essas leituras são realizadas após o processo de perfuração de um poço.

De um modo geral, os dados obtidos com essas duas técnicas — testemunho e perfilagem — têm sido utilizados como as informações mais confiáveis para a identificação das litofácies em poços de petróleo. Dada a existência de muitos campos de exploração, com muitos poços, o banco de dados tende a crescer enormemente, constituindo-se em verdadeiras minas de conhecimento, ‘escondendo’ informações para a identificação das litofácies.

De uma maneira geral, bancos de dados, de diversas áreas, podem ser caracterizados como verdadeiras minas de conhecimento, apenas parcialmente exploradas pelas consultas rotineiras dos usuários [1]. *Mineração de Dados* (MD) é uma área de pesquisa/desenvolvimento que se preocupa em como melhor explorar essas minas. Seu objetivo é descobrir *padrões* (ou *conhecimento*) nos dados. O conhecimento minerado deve ser não trivial, compreensível e de fácil assimilação [2].

A identificação *automática* de litofácies em poços de petróleo pode ser vista como um problema de MD, em que os bancos de dados de perfis e testemunhos são utilizados como minas de conhecimento; com a ajuda de técnicas de MD, o conhecimento para a identificação das litofácies pode ser induzido automaticamente, reduzindo desta forma grande parte do penoso trabalho manual levado a efeito pelos especialistas.

1.1. O Processo de Mineração de Dados

Um processo de MD divide-se em várias etapas — preparação dos dados, amostragem, fragmentação da amostra, indução² de conhecimento e avaliação do conhecimento induzido [3]. O agente do processo é o *minerador*. A seguir detalhamos cada uma dessas etapas:

- *Preparação dos dados* [10] [2] — normalmente, bancos de dados apresentam problemas como dados não informados e dados ‘sujos’. Dados ‘sujos’ são informações corrompidas, e que causam inconsistência a um banco de dados. As causas de dados ‘sujos’ são várias, como por exemplo, erros de digitação. Também nesta fase, pode ser necessário transformar valores contínuos de um atributo em valores discretos. Para ilustrar, considere um atributo com valores contínuos, profundidade de um poço de petróleo: pode ser mais prático trabalhar somente com faixas de valores (profundidade baixa, média e alta).
- *Seleção de amostras* — a importância da seleção de amostras representativas de uma ‘população’ de dados está no provável alto custo de processamento dos algoritmos de indução, se todos os dados disponíveis fossem processados. Como exemplos de técnicas de amostragem, citamos: *Convergence* [10] e *Adaptive Incremental Framework* [11].
- *Fragmentação de amostras* — uma vez escolhida uma amostra, um fragmento dela é utilizado para a indução de conhecimento (próxima etapa) — *conjunto de treinamento* —, enquanto que o outro fragmento é reservado para testar o conhecimento induzido — *conjunto de teste*. São exemplos de técnicas de fragmentação de amostras: *K-fold Cross Validation* [4], *HoldOut* [3] e *Bootstrap* [12].
- *Indução de conhecimento* — nesta etapa é utilizado um dentre os diversos algoritmos de indução para induzir o conhecimento ‘escondido’ no conjunto de treinamento. O conhecimento induzido é apresentado sob um *modelo de conhecimento*. Um modelo de conhecimento é um formalismo de representação de conhecimento, de acordo com o algoritmo de indução empregado. O algoritmo de indução *Prism* [14], por exemplo, induz o conhecimento na forma de regras de classificação *Se <condição> Então <classe>*, em que a classe é um valor de um atributo de classificação (por exemplo, profundidade). Já a família de algoritmos TDIDT (*Top-Down Induction of Decision Trees*) [13] induz conhecimento sob a forma de árvores de decisão. *NaiveBayes* [4], por sua vez, oferece um modelo estatístico. Existem ainda outros

¹ Mais recentemente têm aparecido técnicas de testemunho pós-perfuração, como relatado em [5].

² Neste contexto, induzir e induzir são sinônimos e significam: “deduzir, tirar por conclusão”.

modelos de conhecimento que, por razão de falta de espaço, deixamos de comentar. Modelos baseados em regras de classificação, ou árvores de decisão, entre outros modelos, são genericamente chamados de *classificadores*.

- *Avaliação do conhecimento induzido* [4] — A qualidade de um conhecimento, ou sua confiabilidade, pode ser medida por sua *acurácia*, que é a razão entre o número de testes corretamente classificados e o tamanho do conjunto de teste. Nesta fase, pode haver intensa interação do sistema com o minerador: ele pode, por exemplo, determinar qual é a acurácia mínima exigida para o conhecimento induzido, e tomar as medidas necessárias caso a exigência não tenha sido atendida; para ilustrar, o minerador poderia escolher um outro algoritmo de indução, ou uma nova técnica de fragmentação, etc.

Como vimos, para cada etapa do processo de MD existem diversas técnicas. O problema é que não existe uma técnica que seja considerada a melhor, qualquer que seja a etapa em questão. Mais precisamente, uma combinação de técnicas de amostragem – fragmentação – indução que seja mais adequada para um determinado problema de MD, pode não sê-lo para outro problema. Além da inexistência de técnicas universais, não existem nem mesmo critérios que ajudem o minerador a decidir quais técnicas utilizar em uma dada circunstância.

Retornando ao problema da inferência de conhecimento visando a identificação de litofácies, e abstraindo as técnicas de amostragem, preparação e fragmentação, a dificuldade de escolher o melhor algoritmo de indução de conhecimento vem à tona. Alguns trabalhos pertinentes empregam métodos estatísticos [6]; outros preferem redes neurais artificiais [8], etc. A questão, entretanto permanece: que algoritmo de indução de conhecimento utilizar para induzir o melhor conhecimento possível para a identificação de litofácies?

Em [9], descrevemos um ambiente automatizado para induzir, de bancos de dados da Petrobrás, o conhecimento mais confiável possível para a identificação de litofácies. O ambiente considera a existência de diversas técnicas para cada uma das etapas do processo de MD, escolhendo a melhor combinação dentre as técnicas.

1.2. Organização do Artigo

O objetivo do artigo é apresentar os resultados obtidos com o ambiente automatizado para a identificação de litofácies em reservatórios de petróleo, utilizando para isso os dados do Campo Escola de Namorado da Petrobrás, disponibilizado pela Agência Nacional do Petróleo.

A próxima seção descreve os dados do Campo Escola de Namorado. A seção 3 apresenta os experimentos de MD realizados com dados do Campo Escola de Namorado, assim como os resultados obtidos. A seção 4 é dedicada às conclusões e perspectivas do trabalho.

2. Banco de Dados Campo Escola de Namorado

O Campo Escola de Namorado fica situado na Bacia de Campos no Rio de Janeiro e é composto de 56 poços. O CD-ROM que temos a disposição trás informações sobre os perfis destes 56 poços, dados de testemunhos de 18 poços, além de outros dados.

2.1. Perfis de Poços

Após a perfuração do poço³, normalmente são descidas várias ferramentas visando medir propriedades físicas das formações rochosas. Cada ferramenta é responsável por medir uma propriedade específica. Existem vários tipos de perfis, os mais comumente utilizados são:

- *Potencial Espontâneo (SP)*: é o registro da diferença de potencial entre um eletrodo móvel dentro do poço e outro fixo na superfície. Este perfil permite determinar as camadas permoporosas, calcular a argilidade das rochas, determinar a resistividade da formação e auxilia na correlação com poços vizinhos.
- *Raios Gama (GR)*: permite detectar e avaliar a radioatividade da formação geológica. Utilizado na identificação da litologia, na identificação de minerais radioativos e no cálculo do volume de argila. Também pode ser útil para interpretação de ambientes deposicionais e na investigação da subida do contato óleo/água em reservatórios fraturados.
- *Neutrônico (NPHI)*: as ferramentas mais antigas medem a quantidade de raios gama após a excitação artificial através de bombardeio dirigido de nêutrons rápidos. As ferramentas mais modernas medem a quantidade de nêutrons epitermais e termiais da rocha após o bombardeio. São utilizados para estimativas da porosidade, determinação do volume de argila, identificação da litologia e dos fluidos da formação, além da detecção de hidrocarbonetos ou gás.
- *Indução (ILD)*: fornece uma leitura aproximada da resistividade da formação, através da medição de campos elétricos e magnéticos gerados nas rochas. A resistividade é a propriedade da rocha permitir ou não a passagem de corrente elétrica.

³ Ou de uma fase do poço.

- *Sônico (DT)*: mede a diferença nos tempos de trânsito de uma onda mecânica através das rochas. É utilizada para estimativa da porosidade, identificação da litologia, correlação entre poços, estimativas do grau de compactação das rochas, detecção de fraturas e apoio à sísmica para elaboração de sismogramas.
- *Densidade (RHOB)*: detecta os raios gama defletidos pelos elétrons orbitais dos elementos componentes das rochas, após terem sido emitidos por uma fonte colimada dentro do poço. Além da densidade das camadas, permite o cálculo da porosidade e a identificação de zonas de gás. É utilizado também como apoio à sísmica para o cálculo do sismograma sintético.

Em [7] podem ser encontrados mais detalhes sobre os tipos de perfis aqui mencionados.

~CURVE INFORMATION					
DEPT.M					: Measured Depth
DT.					: 01
GR.					: 02
ILD.					: 03
NPHI.					: 04
RHOB.					: 05
~ASCII LOG DATA					
2952.400	95.4766	74.3750	1.1445	25.7249	2.4929
2952.600	94.8945	78.6992	1.1443	25.7249	2.4480
2952.800	92.9727	80.0625	1.1831	25.7617	2.4355
2953.000	91.8281	81.0391	1.2468	26.0015	2.4119

Figura 1: Perfis de um poço de petróleo.

A figura 1 ilustra os perfis de um dos poços do Campo Escola de Namorado. Os dados encontram-se no formato denominado LAS, que é composto basicamente por um cabeçalho contendo informações sobre o poço e colunas numéricas. Cada coluna numérica representa um dos perfis. A primeira coluna indica a profundidade na qual as propriedades físicas foram medidas. A segunda coluna ilustra o perfil DT, A terceira coluna, o perfil GR. A quarta coluna, o perfil ILD. A quinta coluna, o perfil NPHI. Por fim, a sexta coluna apresenta o perfil RHOB.

Para todos os poços do Campo de Namorado foi feita a leitura dos perfis: GR, ILD, NPHI, RHOB. A leitura do perfil DT só foi realizada em apenas 19 poços.

2.2. Descrição dos Testemunhos

O testemunho é operação especial que pode ocorrer durante a perfuração de um poço de petróleo. O testemunho é o processo de obtenção de uma amostra real da rocha de sub-superfície. Quando um geólogo quer obter uma amostra da formação que está sendo perfurada, a equipe de sonda coloca uma coroa de testemunho na coluna de perfuração. A *coroa de testemunho* é uma broca com um furo no meio que permite o corte do testemunho. O *barrilete de testemunho* é um tubo especial aonde o testemunho irá se alojar e mede geralmente 9, 18 ou 27 metros. O barrilete de testemunho é colocado na parte interna da coluna de perfuração. Durante a perfuração, à medida que a coroa avança, o cilindro de rocha não fragmentado é encamisado pelo barrilete interno e posteriormente trazida a superfície [5].

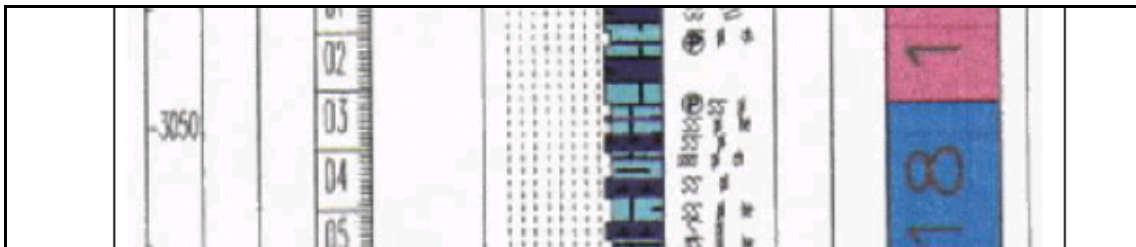


Figura 2: Descrição do testemunho de um poço de petróleo.

Muitas vezes pode haver a necessidade de se testemunhar alguma formação já perfurada. Nesses casos, emprega-se o método de testemunho lateral. Cilindros ocos, presos por cabos 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.

Com a amostra da rocha, os geólogos utilizam seus conhecimentos para determinar, em laboratório, quais as litofácies presentes na amostra, formando a descrição do testemunho.

No Campo Escola de Namorado, apenas 18 poços foram testemunhados. A figura 2 ilustra a descrição do testemunho de um dos poços do campo. A última coluna da figura indica o tipo de litofácies determinada pelo geólogo. Por exemplo, a litofácies 1 representa um *Interlaminação Lamoso Deformado*, já a litofácies 18 indica a litofácies do tipo *Ritmito*. Em [9] podem ser descritas todas as litofácies encontradas no Campo Escola de Namorado.

3. Identificação Automática de Litofácies

Visando induzir o conhecimento necessário para identificar automaticamente as litofácies em poços do Campo Escola de Namorado, unimos os dados disponíveis de testemunho com os dados de perfis de vários poços deste campo. Dessa maneira, foi formado um banco de dados único. Os poços do campo que não possuem testemunho foram descartados, bem como aqueles poços que não contêm todos os cinco perfis — Raios Gama (GR), Sônico (DT), Indução (ILD), Densidade (RHOB) e Porosidade Neutrônica (NPHI).

Como ilustrado na figura 2, os dados de testemunho encontram-se em uma planilha construída manualmente, denominada *descrição de testemunho*, que foram posteriormente digitalizadas. Essa forma de apresentação torna difícil a identificação, em muitos casos, do tipo da litofácies em uma determinada região do testemunho, ou ainda, difícil definir com exatidão a profundidade exata da troca de um tipo de litofácies para outro. Nesses casos, onde o tipo de litofácies não podia ser definido com certeza, os dados foram removidos. A permanência de dados imprecisos no banco de dados para extração de conhecimento poderia dificultar a tarefa do ambiente de extração de conhecimento.

3.1. Resultados

Após a submissão do banco de dados com informações de perfis e testemunhos de poços de petróleo ao ambiente automatizado de MD, foi induzido o conhecimento apresentado parcialmente na figura 3, na forma de um classificador. O classificador induzido obteve 90% de acurácia de teste, isto é, a cada 10 classificações feitas pelo classificador ele acertou 9. Ele é apresentado na forma de uma árvore de decisão, pois foi induzido pelo algoritmo *ID3*⁴.

```

dept.m = [2980,0 - 3016,6]
| ild = [1,0452 - 2,0781]
| | gr = [45,7542 - 49,3945]: litofácies = 17
| | gr = [53,6563 - 57,875]: litofácies = 20
| | gr = [57,875 - 65,3125]: litofácies = 20
| | gr = [65,3125 - 72,2875]
| | | nphi = [20,4688 - 22,6267]: litofácies = 20
...

```

Figura 3: Conhecimento induzido para identificação de litofácies

A interpretação deste classificador pode ser feita da seguinte forma: nas regiões dos poços deste campo onde a profundidade está entre 2980 e 3016,6 metros e o valor do perfil ILD está entre 1,0452 e 2,1781 e o valor do perfil GR está entre 45,7542 e 49,3945, então a litofácies presente é do tipo 17 (Marga Bioturbada). Já caso a profundidade está entre 2980 e 3016,6 metros e o valor do perfil ILD está entre 1,0452 e 2,0781 e o valor do perfil GR está entre 53,5663 e 57,875, a litofácies presente é do tipo 20 (Folhelho com Níveis de Marga Bioturbados) e assim por diante.

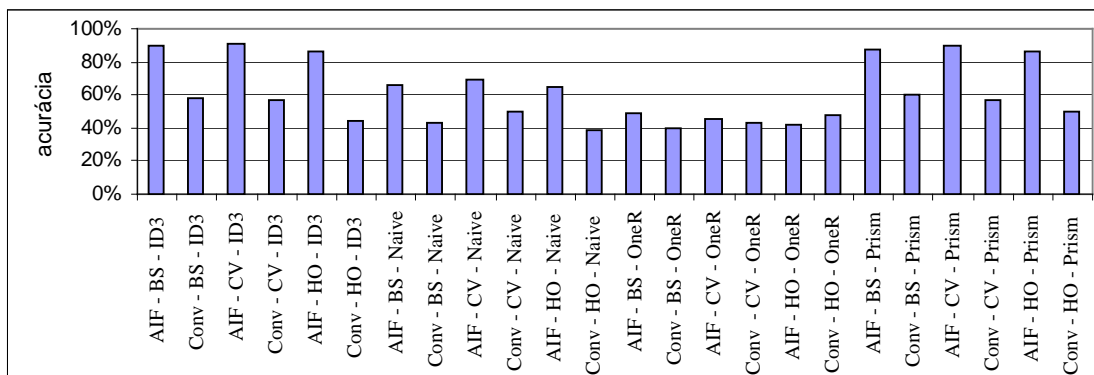


Figura 4: Ranking de classificadores para o banco de dados Campo Escola de Namorado

A figura 4 demonstra os resultados obtidos pelas várias combinações de técnicas de amostragem, fragmentação e indução aplicadas pelo ambiente automatizado de MD. Dentre as técnicas utilizadas, a combinação que obteve melhores resultados foi: *Adaptive Incremental Framework* — amostragem —, *K-fold Cross Validation* — fragmentação

⁴ ID3 é um algoritmo de indução da família TDIDT (*Top-Down Induction of Decision Trees*).

— e *ID3* — indução. Evidencia-se com a figura, a instabilidade no conhecimento induzido com a utilização de combinações de técnicas diferentes para o problema de identificação automática de litofácies. Dado esta instabilidade, o ambiente automatizado de MD obteve bons resultados, induzindo o melhor classificador possível com as técnicas disponíveis — 90% de acurácia média, figura 3.

Além disto, o ambiente automatizado de MD utiliza heurísticas que permitem diminuir o tempo de processamento para a indução do melhor classificador, já que o custo de processamento de todas as combinações possíveis de técnicas pode se tornar inviável. Detalhes sobre as heurísticas estão disponíveis em [9].

4. Conclusões e Perspectivas

Um processo de mineração de dados divide-se em várias etapas — preparação dos dados, amostragem, fragmentação da amostra, indução de conhecimento e avaliação do conhecimento induzido. Para cada etapa, existem diversas técnicas, e nenhuma pode ser considerada melhor que as outras, em todos os casos. O processo deve então ser iterativo, buscando encontrar a combinação de técnicas que conduza ao conhecimento mais confiável.

Neste artigo, discutimos a automação do processo iterativo de mineração de dados, e sua aplicação ao problema de identificação de litofácies presentes em poços de petróleo e gás. Mais precisamente, um classificador de litofácies presentes em poços do campo Escola de Namorado, da Petrobrás, foi induzido automaticamente. Com o classificador, é possível construir facilmente um sistema especialista, para a identificação automática de litofácies de outros poços do campo. A entrada para o sistema especialista seriam dados geológicos de um novo poço; o sistema então classificaria os dados com o classificador induzido, e forneceria como saída à identificação das litofácies presentes no poço. Um sistema assim construído seria de grande valia para os geólogos, diminuindo consideravelmente o tempo gasto pelos mesmos na identificação de litofácies.

Com o processo de mineração automatizado, novos classificadores e sistemas especialistas podem ser construídos, à medida das necessidades. Por exemplo, para um campo de petróleo outro que o Escola de Namorado, um novo classificador seria induzido, utilizando o processo automático de mineração, com o novo classificador sendo à base de um novo sistema especialista.

No estágio atual do trabalho, estamos empenhados em melhorar o rendimento do processo de mineração de dados, adicionando ao processo novas técnicas de preparação de dados, amostragem, fragmentação e indução de conhecimento, bem como pesquisando novas heurísticas, visando diminuir os custos da automação do processo.

5. Referências

- [1] Agrawal, R.; Imielinski T.; Swami A.: “Database Mining: A Performance Perspective”. IEEE Transactions on Knowledge and Data Engineering, Special Issue on Learning and Discovery in Knowledge-Based Databases, Vol. 5, No. 6, December 1993, pp. 914 – 925.
- [2] Fayyad, Usama; Shapiro-Piatetsky, Gregory; Smyth, Padhraic; Uthurusamy, Ramasamy: “Advances in Knowledge Discovery and Data Mining”. AAAI Press, 1996, 611 p.
- [3] Kohavi, Ron: “A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection”. Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 1996, pp. 1137 – 1143.
- [4] Witten, Ian H.; Eibe, Frank: “Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations”. San Diego: Morgan Kaufmann Publishers, 1999, 369 p.
- [5] Thomas, J. E.: “Fundamentos de Engenharia de Petróleo”. Editora: Interciência, 2001.
- [6] Mohn, E.; Berteig, V.; and Helgeland, J.: “A review of statistical approaches to lithofacies determination from well data”. North Sea and Gas Reservoirs. The Norwegian Institute of Technology, pages, 1987, pp. 301–309.
- [7] Doveton, J. H.: “Geologic log interpretation. SEPM Short Course”, No 29. Society for Sedimentary Geology. Tulsa, Oklahoma, 1994.
- [8] Cunha, E. S.; Gomes, H. M.: “Identificação de litofácies de poços de petróleo utilizando um método baseado em redes neurais”. Proceedings of the I Workshop de Teses e Dissertações em Inteligência Artificial, 2002.
- [9] Toebe, Josué; Sampaio, Marcus C.: “Projeto e Construção de um Ambiente para Extrair Conhecimento de Bancos de Dados da Petrobrás”. Dissertação de Mestrado, 2002, 125 p.
- [10] Pyle, Dorian: “Data Preparation for Data Mining”. San Francisco: Morgan Kaufmann Publishers, 1999, 540 p.
- [11] Brumen, B.; Welzer, T.; Jaakkola, H.: “Adaptive Incremental Framework for Performance Driven Data Mining”. Advances in Databases and Information Systems (ADBIS 2001), Lithuania, September 2001, pp. 193 – 203.
- [12] Jain, A. K.; Dubes, R. C.: “Bootstrap Techniques for error Estimation”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, pp. 628 – 633.
- [13] Quinlan, J. R.: “C4.5: Programs for Machine Learning”. San Mateo: Morgan Kaufmann Publishers, 1995.
- [14] Bramer, M. A.: “Automatic Induction of Classification Rules from Examples Using *N-Prism*”. Proceedings in Research and Development in Intelligent Systems XVI, Springer-Verlag, 2000, pp. 99 – 121.
- [15] Thrun, S. B.; Bala, J.; Bloendorn, E.; Bratko, I.: “The Monk’s Problems – A Performance Comparison of Different Learning Algorithms”. Technical Report CMU-CS-91-197, Carnegie Mellon University, 1991.