

## Classificação de Dados Através de Rough Sets

Dante José Alexandre Cid

Marley M. B. R. Vellasco

Marco Aurélio C. Pacheco

Núcleo de Pesquisa em Inteligência Computacional Aplicada

Departamento de Engenharia Elétrica

Pontifícia Universidade Católica do Rio de Janeiro

Rua Marquês de São Vicente, 225

Rio de Janeiro, RJ – Brasil

CEP 22453-900

Tel. / Fax: 55-21-5299445

E-mails: dantecid@ele.puc-rio.br, marley@ele.puc-rio.br, marco@ele.puc-rio.br

### Abstract

*This paper is about application of Rough Sets techniques on Data Classification. In addition to the description of Data Classification task itself, it gives the basics of Rough Sets Theory and details of the application developed to put the concepts of this theory into effect. Two main goals have been achieved: first, the research of a relatively new theory; second, the confirmation of its suitability to the Data Classification task in the context of KDD process.*

### 1. Introdução

A classificação de dados é uma das tarefas mais relevantes no contexto de descoberta de conhecimento em bancos de dados (KDD). Uma das formas mais comuns de se representar o conhecimento extraído de uma base de dados é através de regras de produção, ou regras de inferência, objetivo deste trabalho.

Atualmente, há uma diversidade de técnicas utilizadas para a mineração de dados, algumas delas tradicionais, como os métodos estatísticos, e outras que se inserem na área de Inteligência Computacional, como Algoritmos Genéticos e Redes Neurais. Neste trabalho será abordado o emprego da teoria de Rough Sets para a execução desta tarefa.

A teoria de Rough Sets foi apresentada por Zdzislaw Pawlak no início dos anos 80 como uma abordagem matemática para a análise de dados vagos e imprecisos. O ponto de partida desta teoria é a constatação de que objetos podem ser indiscerníveis (no sentido de similares ou indistinguíveis), devido à limitada informação disponível sobre eles. Neste contexto, objetos que não podem ser especificados através dos dados disponíveis são caracterizados pela teoria de Rough Sets através de dois conceitos precisos:

a aproximação inferior e a aproximação superior, que serão abordadas adiante.

A fundamentação matemática desta teoria permite a descoberta de padrões escondidos na base de dados. Sua utilidade no campo de Mineração de Dados pode ser comprovada pelo crescente número de aplicações e publicações científicas divulgadas com este conteúdo.

Neste artigo o assunto será abordado na seguinte seqüência: após a introdução é feita uma descrição do problema de classificação de dados; em seguida, a descrição da solução adotada para este problema abordará os conceitos da teoria de Rough Sets e detalhes do programa desenvolvido para aplicar estes conceitos; após isto são exibidos os resultados obtidos com a aplicação e a conclusão ao assunto.

### 2. Descrição do Problema

Para melhor entendimento da natureza do problema de classificação de dados, seja considerado o exemplo da **Tabela 1**, representando uma pequena porção de uma base de dados armazenada.

Tabela 1: Informações de lojas de um bairro

LOJA	EXPER. VEND.	QUALID. PROD.	BOA LOCALIZ.	RETOR- NO
1	alta	boa	não	lucro
2	média	boa	não	prejuízo
3	média	boa	não	lucro
4	baixa	média	não	prejuízo
5	média	média	sim	prejuízo
6	alta	média	sim	lucro

As colunas representam os campos ou atributos da informação, e as linhas representam os registros ou instâncias de dados.

Os dados informam características básicas de lojas de um bairro:

- Nível de experiência dos vendedores
- Qualidade do produto vendido
- Se possui boa localização
- Se é lucrativa ou deficitária

Neste caso, o objetivo é realizar uma classificação dos dados acima, obtendo relacionamentos entre conjuntos de atributos preditivos (experiência, qualidade e localização são candidatos) que resultem em determinada classe de um atributo objetivo (no caso, retornar lucro ou prejuízo).

Após o trabalho de mineração poderia ser obtida uma regra como, por exemplo:

“SE EXPERIÊNCIA = alta E QUALIDADE = boa ENTÃO RETORNO = lucro”

Assim, para uma nova loja da qual se deseja prever o retorno, como a de número 7, dadas suas informações:

7	ALTA	BOA	SIM	?
---	------	-----	-----	---

- EXPERIÊNCIA DOS VENDEDORES É ALTA;
- QUALIDADE DO PRODUTO É BOA;
- POSSUI BOA LOCALIZAÇÃO.

Ela poderia ser classificada, com determinada probabilidade de acerto, como lucrativa, pela regra exibida como exemplo.

### 3. Conceitos de Rough Sets

A solução nos casos reais, infelizmente, não é tão simples ou visual como no exemplo dado acima. Para varrer imensos volumes de dados à busca de padrões ocultos, sem mesmo se conhecer que atributos são relevantes para a conclusão a que se deseja chegar, são necessários métodos automáticos e eficazes de busca e análise de informações.

A seguir, será abordada uma das novas técnicas que vêm sendo utilizadas neste sentido: a teoria de Rough Sets, termo que pode ser aproximadamente traduzido como teoria de Conjuntos Imprecisos. Ao longo do artigo será adotado o termo original, por não haver ainda uma consolidação para a melhor tradução para o português.

A teoria de Rough Sets pode ser descrita, de forma bem sumária, como a análise de informações baseada na descoberta de grãos de similaridade existentes no universo dos dados, e no estabelecimento de relações entre estes grãos.

Inicialmente, é importante ressaltar que os conceitos de Rough Sets complementam, não competem com outros métodos, e podem ser usados conjuntamente com outras abordagens como, por exemplo, Lógica Nebulosa, Algoritmos Genéticos, métodos estatísticos, Redes Neurais, etc.

A teoria é de fácil compreensão e aplicação. Diversos sistemas aplicativos baseados em Rough Sets já foram implementados e muitas aplicações em

diversas áreas do conhecimento foram relatadas. Mais detalhes sobre estas aplicações podem ser encontrados em [1].

#### 3.1. Base de Dados e Conceitos Iniciais

Para as considerações acerca dos conceitos, será utilizada como exemplo a Tabela 1, exibida anteriormente.

Naquela pequena base de dados, cada loja é descrita pelo valor dos atributos Experiência, Qualidade, Localização e Retorno, chamados doravante E, Q, L e R, por simplificação.

Cada subconjunto de atributos determina uma partição de todos os objetos (registros) em grãos, compostos por registros que têm o mesmo valor para cada um destes atributos. Por exemplo, os atributos Q e L agregam as lojas nos seguintes grãos: {1,2,3}, {4} e {5,6}. Assim, cada base de dados possui uma família de padrões de classificação que serão usados como base para futuras considerações.

Uma base de dados será formalmente definida da seguinte maneira: um par  $S=(U,A)$ , onde  $U$  e  $A$  são conjuntos finitos, não-vazios, chamados Universo (objetos ou registros) e Atributos, respectivamente.

A cada atributo  $a \in A$  é associado um conjunto  $V_a$  de seus Valores distintos, chamado Domínio de  $a$ . Qualquer subconjunto  $B$  de  $A$  determina uma relação binária  $I(B)$  em  $U \times U$ , chamada de Relação de Indiscernibilidade, definida a seguir:

$(x,y) \in I(B)$  se e somente se  $a(x) = a(y)$  para todo  $a \in B$ , onde  $a(x)$  é o valor do atributo  $a$  no elemento (objeto ou registro)  $x$ .

Em [2] é demonstrado que  $I(B)$  é uma relação de equivalência. A família de todas as classes de  $I(B)$  é a partição determinada por  $B$  em  $U$ , e é denominada  $U/I(B)$ , ou simplesmente  $U/B$ . Uma classe de equivalência de  $I(B)$  (um bloco da partição) contendo  $x$  será identificado como  $B(x)$ .

Se  $(x,y)$  pertence a  $I(B)$  diz-se que  $x$  e  $y$  são  $B$ -indiscerníveis. As classes de equivalência da relação  $I(B)$  são referidas como Conjuntos Elementares por  $B$  ou grãos- $B$ .

A relação de equivalência não é suficiente como base para a teoria de Rough Sets. Relações de tolerância e de ordenação, entre outras, são propostas, por exemplo, em [3,4]. Para o escopo deste trabalho, porém, será suficiente como base a relação de equivalência, tendo em vista não tornar demasiadamente complexo o primeiro contato com a teoria.

#### 3.2. Aproximações

Inicialmente seja considerada a seguinte questão: Quais são as características que definem uma loja como dando lucro ou prejuízo, a partir das informações da

Tabela 1? Pode-se notar que não há uma resposta única para esta pergunta, pois as lojas 2 e 3 têm as mesmas características pelos atributos E, Q e L, mas a loja 2 é lucrativa enquanto a 3, não. Em vista dos mesmos atributos, pode-se dizer com certeza que, no universo em questão, qualquer loja com valores de E, Q e L iguais aos das lojas 1 e 6 dão lucro, assim como qualquer loja com valores de E, Q e L iguais aos das lojas 4 e 5 dão prejuízo, mas nada se pode afirmar para lojas com valores de E, Q e L iguais aos das lojas 2 e 3.

Dado o conjunto  $\mathbf{B} = \{E, Q, L\}$ , um grão-B é um conjunto de registros (no caso, lojas) que possuem o mesmo valor para os atributos de B. Por exemplo, o grão-B caracterizado por  $(E, Q, L) = (\text{média}, \text{boa}, \text{não})$  é representado pelo conjunto de lojas  $\{2,3\}$ . Já o grão-B caracterizado por  $(E, Q, L) = (\text{alta}, \text{média}, \text{sim})$  é representado pelo conjunto unitário  $\{5\}$ .

A Aproximação Inferior é formada pela união dos grãos que estão completamente contidos no conceito, isto é, grãos em que não há elementos (lojas) de classe diferente de lucro. De acordo com a Tabela 1, para  $\mathbf{X}=\{1,3,6\}$  e  $\mathbf{B}=\{E, Q, L\}$ , somente os dois grãos-B representados por  $\{1\}$  e  $\{6\}$  satisfazem a esta definição. Portanto, o conjunto  $\{1,6\}$  formado pela união dos grãos  $\{1\}$  e  $\{6\}$  compõe a chamada Aproximação Inferior ou Região Positiva de  $\mathbf{X}$  por  $\mathbf{B}$ .

A Aproximação Superior é formada pela união dos grãos que têm interseção não nula com o conceito, isto é, grãos em que há elementos (lojas) de classe igual a lucro. De acordo com a Tabela 1, para  $\mathbf{X}=\{1,3,6\}$  e  $\mathbf{B}=\{E, Q, L\}$ , os grãos-B representados por  $\{1\}$ ,  $\{2,3\}$  e  $\{6\}$  satisfazem a esta definição. Portanto, o conjunto  $\{1,2,3,6\}$  formado pela união dos grãos  $\{1\}$ ,  $\{2,3\}$  e  $\{6\}$  compõe a chamada Aproximação Superior de  $\mathbf{X}$  por  $\mathbf{B}$ .

O conjunto formado pelos grãos da Aproximação Superior que não pertencem à interseção desta com a Aproximação Inferior compõe o que é denominado Região de Fronteira. No exemplo dado, o conjunto  $\{2,3\}$  forma a Região de Fronteira de  $\mathbf{X}$  por  $\mathbf{B}$ .

As aproximações podem ser definidas formalmente como dois conjuntos,  $B_*(X)$  e  $B^*(X)$ , chamados respectivamente de Aproximação Inferior de  $\mathbf{X}$  por  $\mathbf{B}$  e Aproximação Superior de  $\mathbf{X}$  por  $\mathbf{B}$ , conforme a definição a seguir:

O conjunto denominado Região de Fronteira de  $\mathbf{X}$  por  $\mathbf{B}$  é definido como:

$$BN_B(X) = B^*(X) - B_*(X)$$

Se a região de fronteira for o conjunto vazio, isto é,  $BN_B(X) = \emptyset$ , então  $X$  é crisp (preciso) com relação a  $\mathbf{B}$ .

Caso contrário, isto é,  $BN_B(X) \neq \emptyset$ , então  $X$  é dito

$$B_*(X) = \bigcup_{x \in U} \{B(x) : B(x) \subseteq X\}$$

$$B^*(X) = \bigcup_{x \in U} \{B(x) : B(x) \cap X \neq \emptyset\}$$

rough (impreciso) com relação a  $\mathbf{B}$ .

A **Figura 1** exibe uma forma gráfica de representar o conceito das aproximações.

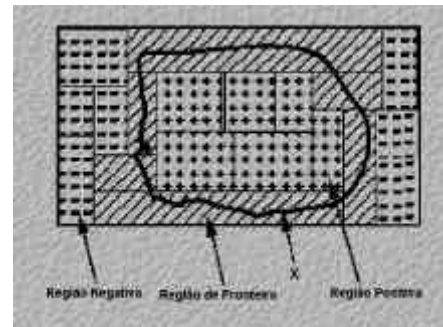


Figura 1: As aproximações em forma gráfica.

O Grau de Imprecisão, ou Roughness, de um conjunto pode ser caracterizado numericamente como:

$$\alpha_B(X) = \left( \frac{\text{card}(B_*(X))}{\text{card}(B^*(X))} \right)$$

Se  $\alpha_B(X) = 1$ , então  $X$  é **preciso** (*crisp*) com relação a  $\mathbf{B}$ . Senão,  $X$  é **impreciso** (*rough*) com relação a  $\mathbf{B}$ . Por exemplo, o Grau de Imprecisão de  $X = \{1,3,6\}$  é dado por  $\text{card}(\{1,6\}) / \text{card}(\{1,2,3,6\}) = 2/4 = 0,5$ .

### 3.3. Dependência de Atributos

Outra questão importante na análise de dados é descobrir dependências entre atributos. Suponha que um conjunto  $\mathbf{A}$  de atributos de uma base de dados  $\mathbf{S}=(U, \mathbf{A})$  seja dividido em dois subconjuntos  $\mathbf{P}$  e  $\mathbf{O}$ , chamados atributos **preditivos** e **objetivos**, respectivamente, tal que  $\mathbf{P} \dot{\cup} \mathbf{O} = \mathbf{A}$  e  $\mathbf{P} \cap \mathbf{O} = \emptyset$ . Tal base de dados é chamada *tabela de decisão*.

Intuitivamente, um conjunto de atributos  $\mathbf{O}$  depende totalmente de  $\mathbf{P}$  ( $\mathbf{P} \dot{\supset} \mathbf{O}$ ), se todos os valores de  $\mathbf{O}$  forem univocamente determinados por valores de atributos de  $\mathbf{P}$ . Em outras palavras,  $\mathbf{O}$  depende totalmente de  $\mathbf{P}$  se existe uma dependência funcional entre valores de  $\mathbf{P}$  e  $\mathbf{O}$ .

Formalmente, a dependência pode ser assim definida:

Sejam  $\mathbf{P}$  e  $\mathbf{O}$  subconjuntos de  $\mathbf{A}$ . O Grau de Dependência de  $\mathbf{O}$  em relação a  $\mathbf{P}$  é dado por:

$$g(\mathbf{P}, \mathbf{O}) = \frac{\text{card}(P_*(X))}{\text{card}(U)}$$

Se  $g(\mathbf{P}, \mathbf{O}) = 1$  diz-se que  $\mathbf{O}$  depende totalmente de  $\mathbf{P}$ ; se  $g(\mathbf{P}, \mathbf{O}) = 0$ ,  $\mathbf{O}$  não depende de  $\mathbf{P}$ ; se  $0 < g(\mathbf{P}, \mathbf{O}) < 1$ ,  $\mathbf{O}$  depende parcialmente de  $\mathbf{P}$ ;

Por exemplo, na **Tabela 1** o grau de dependência de  $O=\{R\}$  em relação a  $P=\{E,Q,L\}$  é dado por  $\text{card}(\{1,6\}) / \text{card}(\{1,3,6\}) = 2/3$ .

### 3.4 Redução de Atributos

Uma redução é o conjunto mínimo de atributos que preserva o grau de dependência. Em outras palavras, uma redução é o subconjunto mínimo de atributos preditivos que permite tomar as mesmas decisões do conjunto completo.

Por exemplo, na **Tabela 1** existem duas reduções,  $\{E,Q\}$  e  $\{E,L\}$ , dos atributos preditivos  $\{E,Q,L\}$ .

A redução de atributos é uma questão básica na teoria de Rough Sets [2]. O cálculo de reduções é um problema n-p completo, e seu processamento em grandes bases de dados exige grande esforço computacional.

### 3.5 Significância de Atributos

O conceito de redução nos leva a desconsiderar alguns atributos de tal forma que as relações básicas no banco de dados sejam preservadas. Alguns atributos, no entanto, têm maior relevância que outros na preservação destas relações. Este conceito é chamado de Significância do atributo, e pode ser assim definido:

Sejam  $\mathbf{P}$  e  $\mathbf{O}$  conjuntos de atributos preditivos e objetivos, respectivamente, e seja  $\mathbf{a}$  um atributo preditivo pertencente a  $\mathbf{P}$ . A significância de  $\mathbf{a}$  será calculada em função da mudança do Grau de Dependência de  $\mathbf{O}$  em relação a  $\mathbf{P}$  com a remoção de  $\mathbf{a}$ , segundo a fórmula abaixo:

$$S_{(P,O)}(a) = 1 - \left( \frac{g(P - \{a\}, O)}{g(P, O)} \right)$$

Este coeficiente pode ser visto como o erro que ocorre na definição de  $\mathbf{O}$  por  $\mathbf{P}$  quando  $\mathbf{a}$  é removido. Por exemplo, para  $X = \{1,3,6\}$ ,  $O=\{R\}$ ,  $P=\{E,Q,L\}$  e  $\mathbf{a} = \{L\}$ , o Grau de Significância do atributo  $\mathbf{a}$  é dado por  $1 - (\gamma(\{E,Q\}) / \gamma(\{E,Q,L\})) = 1 - ((2/3) / (2/3)) = 0$ , significando que remover o atributo L de consideração não afetará os resultados. O mesmo resultado seria encontrado ao se remover Q de  $\{E,Q,L\}$ .

### 3.6 Regras de Decisão

As dependências entre atributos encontradas com as técnicas de Rough Sets descritas anteriormente podem ser expressas na forma de regras de decisão. Este é o ponto de ligação entre a teoria e a aplicação em mineração de dados a que este trabalho se propõe.

Para exemplificar regras geradas pelos métodos abordados, seja a redução  $P=\{E,Q\}$  da base da **Tabela**

1. Algumas regras de decisão descrevendo a dependência de  $\{R\}$  em relação a P poderiam ser:

- Se E = ALTA e Q = BOA ENTÃO R = LUCRO
- Se E = MÉDIA e Q=BOA ENTÃO R=PREJUÍZO
- Se E = MÉDIA e Q = BOA ENTÃO R = LUCRO
- Se E=BAIXA e Q=MÉDIA ENTÃO R=PREJUÍZO
- Se E = ALTA e Q = MÉDIA ENTÃO R = LUCRO

Pode-se notar, sob a ótica dos dados fornecidos, que algumas regras extraídas são 100% válidas, enquanto outras têm um nível de confiabilidade menor.

Um conjunto de regras de decisão normalmente é chamado de base de conhecimento. A cada regra podem ser associados dois coeficientes básicos que denotam sua qualidade: o Fator de Certeza (também chamado de eficácia) e o Fator de Cobertura (também chamado de abrangência), definidos a seguir.

Seja a regra de decisão “SE p ENTÃO q”. O Fator de Certeza desta regra é dado por:

$$P(q/p) = \left( \frac{\text{card}(p \wedge q)}{\text{card}(p)} \right)$$

O Fator de Certeza indica o quanto a regra acerta a previsão, dadas suas pré-condições  $\mathbf{p}$ . Tomando como exemplo a regra “Se E = MÉDIA e Q = BOA ENTÃO R = LUCRO”, tem-se  $\mathbf{p} = \{E = MÉDIA \text{ e } Q = BOA\}$  e  $\mathbf{q} = \{R = LUCRO\}$ . As lojas 2 e 3 satisfazem  $\mathbf{p}$ , enquanto somente a loja 3 satisfaz  $\mathbf{p}$  e  $\mathbf{q}$ . O Fator de Certeza desta regra é, portanto,  $1/2 = 50\%$ .

Dada a regra “SE p ENTÃO q”, seu Fator de Cobertura é dado por:

$$P(q/p) = \left( \frac{\text{card}(p \wedge q)}{\text{card}(q)} \right)$$

O Fator de Cobertura indica o quanto a regra abrange os casos em que o atributo objetivo tem um determinado valor. No exemplo da regra “Se E = MÉDIA e Q = BOA ENTÃO R = LUCRO”,  $\mathbf{p} = \{E = MÉDIA \text{ e } Q = BOA\}$  e  $\mathbf{q} = \{R = LUCRO\}$ . As lojas 1, 3 e 6 satisfazem  $\mathbf{q}$ , enquanto somente a loja 3 satisfaz  $\mathbf{p}$  e  $\mathbf{q}$ . O Fator de Cobertura desta regra é, portanto,  $1/3 = 33,3\%$ .

## 4. Aplicativo desenvolvido

A partir da teoria exposta até aqui, foi idealizada a construção de um sistema implementando os conceitos de Rough Sets para extrair regras de conhecimento de uma base de dados.

O sistema foi modelado segundo o diagrama de blocos da **Figura 2**.

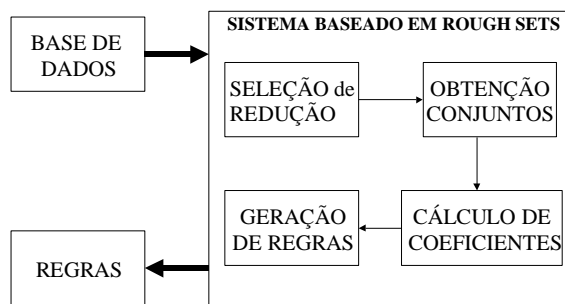


Figura 2: Diagrama de blocos do sistema

O processo de redução, conforme mencionado anteriormente, é um problema à parte no processo de mineração de dados e, neste sentido, está além do escopo do trabalho. Portanto, para avaliar as diversas opções de conjuntos reduzidos de atributos, foi gerada uma tabela contendo, em cada registro, uma diferente opção de redução.

O sistema acessa a base de dados escolhida e, para cada redução selecionada, realiza operações que particionam a base segundo as aproximações da teoria. A partir dos conjuntos definidos pelas aproximações são feitos os cálculos dos diversos coeficientes detalhados anteriormente. A partir deles, são selecionados os elementos mais representativos para compor uma regra de decisão. A regra é então criada a partir do conteúdo do elemento e em seguida adicionada ao banco de regras.

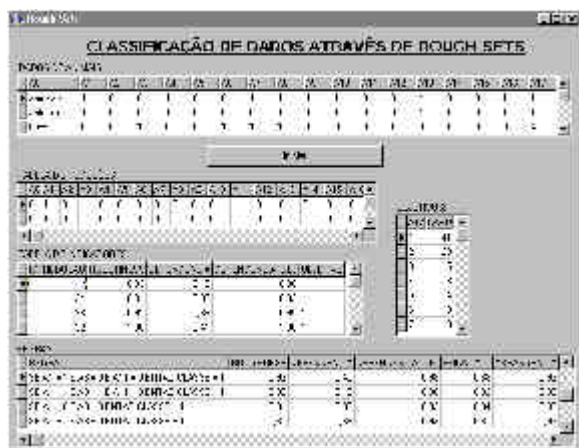


Figura 3 – Tela principal do sistema

A Figura 3 exibe a tela principal do sistema. Por tratar-se de uma aplicação didática, a principal preocupação foi com a exibição de todos os dados envolvidos em cada etapa da execução, desde a base de dados até a regra gerada, passando pelas reduções, conjuntos e coeficientes.

A linguagem escolhida para implementação foi o C++. Quanto à base de dados, a opção recaiu sobre o Borland Interbase, pela boa relação entre rapidez e facilidade de instalação e a capacidade do SQL nativo.

A base de dados escolhida para avaliação foi retirada do site de acesso público em: “ftp://ftp.ics.uci.edu/pub/machine-learning-databases”. Foi escolhido o banco chamado “zoo”, que contém 18 atributos de 101 diferentes animais, classificados em sete categorias. A base foi dividida entre treinamento e teste a 50% de cada classe.

## 5. Resultados

Após a execução do programa, foi examinada a tabela de regras gerada e avaliados seus coeficientes. A Tabela 2 resume os resultados encontrados para algumas das principais regras, testadas sobre a base de treinamento. Foram colocados os resultados para a quantidade mínima de regras em cada classe que abrangesse a totalidade de seus registros. Por isso, algumas classes têm mais de uma regra exibida.

As regras foram em seguida aplicadas sobre a base de teste. Os resultados são exibidos na Tabela 3, onde se pode notar que foram colocadas as mesmas regras da Tabela 2. O motivo foi que nem sempre um mesmo conjunto de regras conseguia abranger todos os elementos de uma classe com boa eficácia. Assim, há regras na Tabela 2 que não representam bem o conjunto de treinamento, mas foram muito boas para o conjunto de teste. Consideração análoga vale para a Tabela 3. A Tabela 4 mostra o texto das regras indicadas.

Tabela 2 – Resultados das regras sobre a base de treinamento

REGRA	CLASSE	EFICÁCIA	ABRANGÊNCIA	REGISTROS NA CLASSE
1	1	100%	100%	20
118	2	100%	100%	10
202	3	100%	100%	2
211	3	100%	50%	2
237	3	11%	50%	2
282	4	100%	100%	6
351	5	100%	100%	2
408	6	100%	100%	4
409	6	100%	100%	4
478	7	100%	100%	5
488	7	71%	100%	5

Tabela 3 – Resultados das regras sobre a base de teste

REGRA	CLASSE	EFICÁCIA	ABRANGÊNCIA	REGs NA CLASSE
1	1	100%	100%	21
118	2	100%	100%	10
202	3	0%	0%	3
211	3	100%	67%	3
237	3	33%	100%	3
282	4	100%	100%	7
351	5	67%	100%	2
408	6	60%	75%	4
409	6	100%	100%	4
478	7	100%	20%	5
488	7	71%	100%	5

Tabela 4 – Texto das regras

Nº REGRA	TEXTO DA REGRA
1	SE milk = 1 ENTÃO type = 1
118	SE milk = 0 E legs = 2 ENTÃO type = 2
202	SE hair = 0 E venomous = 1 ENTÃO type = 3
211	SE eggs = 1 E aquatic = 0 E toothed = 1 ENTÃO type = 3
237	SE eggs = 1 E aquatic = 0 E backbone = 1 ENTÃO type = 3
282	SE breathes = 0 E fins = 1 ENTÃO type = 4
351	SE aquatic = 1 E toothed = 1 E legs = 4 ENTÃO type = 5
408	SE feathers = 0 E aquatic = 0 E predator = 0 E backbone = 0 ENTÃO type = 6
409	SE aquatic = 0 E toothed = 0 E legs = 6 ENTÃO type = 6
478	SE breathes = 0 E venomous = 0 E tail = 0 ENTÃO type = 7
488	SE hair = 0 E toothed = 0 E backbone = 0 ENTÃO type = 7

Como pode ser visto, foram obtidas regras de elevada qualidade. Porém, a pequena quantidade de registros para algumas classes nesta base trazem efeitos negativos à tentativa de generalização do conhecimento por regras. Uma classe como a 6, por exemplo, torna difícil a obtenção de coeficientes elevados, pois cada erro cometido faz o coeficiente de abrangência cair em 25%.

A fim de se obter uma medida de comparação entre os resultados alcançados com técnicas de Rough Sets e outros métodos, a mesma base de dados foi submetida a um sistema Neuro-Fuzzy-Hierárquico (híbrido entre Redes Neurais e Lógica Fuzzy), chamado NFHB, desenvolvido também no ICA/PUC-RJ para classificação de dados. A mesma divisão entre conjuntos de treinamento e teste foi respeitada e os resultados obtidos são comparados na **Tabela 5**, através da eficácia e abrangência das principais regras que classificam todas as instâncias de cada classe. Pode-se notar que, para cada classe, cada sistema pode necessitar de uma quantidade diferente de regras para representar todos os elementos. É importante também ressaltar que em cada linha são colocadas as melhores regras de cada sistema, que não são necessariamente iguais.

Tabela 5 – Comparação de resultados

CLASSE	NFHB		ROUGH SETS	
	EFICÁCIA	ABRANGÊNCIA	EFICÁCIA	ABRANGÊNCIA
1	99,7	38,5	100	100
2	90,5	13,6	100	100
2	93	1,4	-	-
3	62,3	2,4	100	67
3	-	-	33	100
4	98,7	8,1	100	100
5	46,9	4,1	67	100
6	91,8	4	100	100
6	99,9	3,4	-	-
7	94,6	8,2	100	20
7	-	-	71	100

## 6. Conclusão

Além de atingir o objetivo de estudar e divulgar a teoria de Rough Sets, relativamente nova mas já de considerável relevo na comunidade internacional, este trabalho comprovou, através da aplicação desenvolvida, a adequação dos conceitos à tarefa de classificação de dados.

A mineração de dados é uma questão bastante relevante nos dias atuais, e muitos métodos têm sido propostos para as diversas tarefas que dizem respeito a esta questão. Sem dúvida, a teoria de Rough Sets mostrou-se de grande valia na busca de novos caminhos para a descoberta de conhecimento. Podem ser destacadas algumas das principais vantagens do seu uso neste contexto [2]:

- ◆ Provê eficientes algoritmos para a descoberta de padrões ocultos nos dados.
- ◆ Indica o caminho para reduções
- ◆ Avalia a significância de dados
- ◆ Gera conjunto mínimo otimizado de regras
- ◆ É de fácil compreensão
- ◆ Oferece resultados facilmente interpretáveis

Muitas possibilidades se abrem para trabalhos futuros, particularmente no sentido de conjugar as técnicas de Rough Sets com outras, como Redes Neurais, Lógica Nebulosa e Algoritmos Genéticos, principalmente na tarefa de redução de atributos.

## Bibliografia

- [1] Polkowski L., Skowron A. - Rough Sets in Knowledge Discovery, Methodology and Applications, Vol. 1-2, Physica-Verlag.
- [2] Pawlak, Z. - Data Mining – A Rough Set Perspective, Polish Academy of Sciences.
- [3] Komorowski J., Polkowski L., Skowron A. - Rough Sets: A Tutorial, Institute of Mathematics, Warsaw University.
- [4] Stefanowski J. - On Rough Set Based Approaches to Induction of Decision Rules, Rough Sets in Knowledge Discovery Vol. 1, Physica-Verlag.