

Método para Geração de Regras de Classificação Não-Determinística Baseado em Rough Sets

C.M.M.M. PATRÍCIO¹, J.O.P. PINTO², Laboratório de Inteligência Artificial, Eletrônica de Potência e Digital - BATLAB, Universidade Federal de Mato Grosso do Sul - UFMS, Cidade Universitária s/n, Cx.P. 549, 79070-900 Campo Grande, MS, Brasil

P.P. da SILVA³, Agência Estadual de Regulação de Serviços Públicos de Mato Grosso do Sul - AGEPAN, Rua Barão do Rio Branco, 1092, 79002-175 Campo Grande, MS, Brasil.

Resumo. O objetivo desse trabalho é apresentar um método baseado em Rough Sets, capaz de gerar regras de classificação não-determinística, que permite ao usuário especificar o mínimo de consistência que a regra de classificação terá que satisfazer e somente gerar regras que atendam a este requisito. Isto quer dizer que, neste método, as regras são requeridas para dar suficiente suporte e serem consistentes somente o necessário ao banco de dados.

1. Introdução

Com o grande avanço tecnológico que o mundo vem alcançando nos últimos anos, disponibilizou-se diversos recursos para o armazenamento de dados. Com isso os sistemas de informações formados através dos bancos de dados, na sua maioria, tornaram-se desnecessariamente grandes, pois nem todas as informações armazenadas são utilizadas pelos usuários para qualquer processo de tomada de decisão, principalmente gerencial. Em 1982, Pawlak [4] introduziu a Teoria de Rough Sets que é utilizada para a mineração e obtenção de conhecimento em banco de dados. Diversos estudos estão sendo realizados nos últimos anos com esta teoria que tem se revelado como sendo uma poderosa ferramenta para mineração de dados e para a descoberta de conhecimento em banco de dados [5].

Em [2], por exemplo, mostra-se como a teoria de Rough Sets possui propriedades que permitem eliminar atributos irrelevantes através do processo de redução do sistema de informação, baseando-se na definição de redutos, os quais são subconjuntos de atributos capazes de manter as mesmas propriedades da representação de conhecimento quando esta é feita utilizando todos os atributos. A redução do sistema de informação, e conseqüentemente a geração de regras, são os objetivos desta teoria

¹cristian_mara@mail.uniderp.br

²jpinto@del.ufms.br

³psilva@agepan.ms.org.br, também atua no BATLAB, UFMS

uma vez que o sistema de informação pode ser desnecessariamente grande. Esta redução implica diretamente na diminuição do custo computacional do sistema.

CID [1], em seu trabalho, avalia o desempenho da técnica de Rough Sets na tarefa de Classificação de Dados e comprova a adequação dos conceitos desta teoria a esta aplicabilidade. Relata ainda, alguns pontos frágeis da técnica que foram identificados, como a necessidade de um mecanismo de apoio para a redução de atributos e a dificuldade em trabalhar com atributos de domínio contínuo.

Diversos métodos baseados nessa teoria foram desenvolvidos para conseguir regras de classificação [7, 6]. Alguns destes métodos são determinísticos, isto é, deles derivam-se regras de classificação totalmente consistentes com os dados do banco de dados. Enquanto esta característica determinística é desejável em algumas aplicações, ela apresenta dois problemas como método de classificação geral:

- Primeiramente, ela não pode trabalhar com incerteza de dados de forma efetiva. Ela não considera uma classificação que é 99% consistente com o banco de dados e considera o 1% restante como ruído de dados. Isto pode fazer com que algumas valiosas informações deixem de ser descobertas, por serem consideradas ruídos.
- Em segundo lugar algumas regras de classificação satisfarão um pequeno número de objetos no banco de dados.

A parte mais difícil na classificação de banco de dados é encontrar um número mínimo de regras que caracterizem cada um dos objetos de um sistema de decisão. Em [3] apresenta-se a teoria de Rough Sets, capaz de gerar regras de classificação determinística. Será apresentado, neste artigo, um método baseado em Rough Sets, capaz de gerar regras de classificação não-determinística, que permite ao usuário especificar o mínimo de consistência que a regra de classificação terá que satisfazer e somente gerar regras que atendam a este requisito. Isto quer dizer que, neste método, as regras são requeridas para dar suficiente suporte e serem consistentes somente o necessário ao banco de dados.

2. Classificação Rough

Considerando a seguinte situação: uma empresa irá disponibilizar, para comercialização no mercado, brinquedos infantis. Ela necessita saber, antes de qualquer coisa, qual será a aceitação ou rejeição das crianças, seu público alvo, em relação aos seus brinquedos. Para tanto, disponibilizou diversos brinquedos com características diferentes, e selecionou-se dez crianças para realizarem um teste. A pergunta é: qual a atitude das crianças com relação aos brinquedos? [3].

A forma mais comum para a representação dos dados na abordagem de Rough Sets é através de um sistema de decisão que contém um conjunto de objetos, sendo que cada objeto tem uma quantidade de atributos condicionais e um conjunto de atributo de decisão. Estes atributos condicionais e de decisão são os mesmos para cada um dos objetos, mas seus valores nominais podem diferir. O sistema de decisão capaz de fornecer a atitude das crianças com relação aos brinquedos será representado por \mathfrak{R} e encontra-se na Tabela 1, sendo U o conjunto de objetos, C o conjunto

dos atributos condicionais disponíveis na base de dados e D o conjunto do atributo de decisão. Assim, o sistema de decisão é dado por $\mathfrak{R} = (U, C \cup D)$.

Tabela 1: Sistema de Decisão - \mathfrak{R} .

U	<i>Atributos Condicionais (C)</i>					<i>Atributo de Decisão (D)</i>
Criança	Cor	Tamanho	Tato	Textura	Material	Atitude
1	<i>Azul</i>	<i>Grande</i>	<i>Duro</i>	<i>Indefinido</i>	<i>Plástico</i>	<i>Negativa</i>
2	<i>Vermelho</i>	<i>Médio</i>	<i>Moderado</i>	<i>Liso</i>	<i>Madeira</i>	<i>Neutra</i>
3	<i>Amarelo</i>	<i>Pequeno</i>	<i>Macio</i>	<i>Áspero</i>	<i>Pelúcia</i>	<i>Positiva</i>
4	<i>Azul</i>	<i>Médio</i>	<i>Moderado</i>	<i>Áspero</i>	<i>Plástico</i>	<i>Negativa</i>
5	<i>Amarelo</i>	<i>Pequeno</i>	<i>Macio</i>	<i>Indefinido</i>	<i>Plástico</i>	<i>Neutra</i>
6	<i>Verde</i>	<i>Grande</i>	<i>Duro</i>	<i>Liso</i>	<i>Madeira</i>	<i>Positiva</i>
7	<i>Amarelo</i>	<i>Pequeno</i>	<i>Duro</i>	<i>Indefinido</i>	<i>Metal</i>	<i>Positiva</i>
8	<i>Amarelo</i>	<i>Pequeno</i>	<i>Duro</i>	<i>Indefinido</i>	<i>Plástico</i>	<i>Positiva</i>
9	<i>Verde</i>	<i>Grande</i>	<i>Duro</i>	<i>Liso</i>	<i>Madeira</i>	<i>Neutra</i>
10	<i>Verde</i>	<i>Médio</i>	<i>Moderado</i>	<i>Liso</i>	<i>Plástico</i>	<i>Neutra</i>

Para a simplificação das tabelas, será considerado as abreviações conforme a Tabela 2.

Tabela 2: Abreviações Consideradas.

<i>Atributos Condicionais (C)</i>	<i>Valores Nominais</i>
Cor = Co	<i>Azul = Az; Vermelho = Vm; Amarelo = Am; Verde = Vd</i>
Tamanho = Ta	<i>Grande = Gr; Médio = Me; Pequeno = Pq</i>
Tato = To	<i>Duro = Du; Moderado = Mo; Macio = Mc</i>
Textura = Te	<i>Indefinido = In; Liso = Li; Áspero = As</i>
Material = Ma	<i>Plástico = Pl; Pelúcia = Pe; Madeira = Md; Metal = Mt</i>
<i>Atributo de Decisão (D)</i>	<i>Valores Nominais</i>
Atitude = At	<i>Positiva = Po; Negativa = Ng; Neutra = Nt</i>

Neste trabalho, é acrescentado um parâmetro ao método Rough Sets apresentado por [3], de modo a permitir que as regras de classificação não-determinísticas sejam obtidas. Para melhor entendimento, será utilizado o sistema de decisão representado por \mathfrak{R} , conforme Tabela 3. Todos os atributos de \mathfrak{R} já são conhecidos com exceção da última coluna que é usada para indicar o Número de Registros Iguais - NRI, ocorridos no sistema de decisão. No caso desta tabela, não existe registros iguais para este sistema, portanto todas as linhas da coluna NRI apresentam valor unitário.

A coluna NRI inserida no sistema \mathfrak{R} é de fundamental importância para este método de geração de regras não-determinísticas, pois ela é usada para obter in-

Tabela 3: Sistema de Decisão - **R**.

<i>U</i>	<i>Atributos Condicionais (C)</i>					<i>Atributo de Decisão (D)</i>	<i>Número Reg. Iguais</i>
Criança	Co	Ta	To	Te	Ma	At	NRI
1	<i>Az</i>	<i>Gr</i>	<i>Du</i>	<i>In</i>	<i>Pl</i>	<i>Ng</i>	1
2	<i>Vm</i>	<i>Me</i>	<i>Mo</i>	<i>Li</i>	<i>Md</i>	<i>Nt</i>	1
3	<i>Am</i>	<i>Pq</i>	<i>Mc</i>	<i>As</i>	<i>Pe</i>	<i>Po</i>	1
4	<i>Az</i>	<i>Me</i>	<i>Mo</i>	<i>As</i>	<i>Pl</i>	<i>Ng</i>	1
5	<i>Am</i>	<i>Pq</i>	<i>Mc</i>	<i>In</i>	<i>Pl</i>	<i>Nt</i>	1
6	<i>Vd</i>	<i>Gr</i>	<i>Du</i>	<i>Li</i>	<i>Md</i>	<i>Po</i>	1
7	<i>Am</i>	<i>Pq</i>	<i>Du</i>	<i>In</i>	<i>Mt</i>	<i>Po</i>	1
8	<i>Am</i>	<i>Pq</i>	<i>Du</i>	<i>In</i>	<i>Pl</i>	<i>Po</i>	1
9	<i>Vd</i>	<i>Gr</i>	<i>Du</i>	<i>Li</i>	<i>Md</i>	<i>Nt</i>	1
10	<i>Vd</i>	<i>Me</i>	<i>Mo</i>	<i>Li</i>	<i>Pl</i>	<i>Nt</i>	1

formações das regras quanto a dois parâmetros inseridos no processo de classificação: suporte e consistência.

Dada uma regra $r \rightarrow$ “SE *atributos condicionais* ENTÃO *atributo de decisão*”, a partir dessa regra pode-se definir o suporte e a consistência da seguinte maneira:

- Suporte (φ) - O suporte da regra é definido como o número de objetos em **R** que tenham os mesmos valores nominais dos atributos condicionais e do atributo de decisão de r .
- Consistência (δ) - A consistência da regra r é definida pela razão entre o número de objetos em **R** que tenham os mesmos valores nominais dos atributos condicionais e do atributo de decisão, ou seja, o suporte de r , e o número de objetos em **R** que tenham os mesmos valores nominais dos atributos condicionais.

Por exemplo: considerando a regra $r \rightarrow$ “SE ($Co = Vd$ e $Ta = Gr$ e $To = Du$ e $Te = Li$ e $Ma = Md$) ENTÃO ($At = Po$)”, do sistema de decisão da Tabela 3, o suporte e consistência para esta regra será:

- Suporte: $\varphi = 1$ (referente ao objeto 6)
- Consistência: $\delta = \varphi/2 = 0,5$ (referente aos objetos 6 e 9).

Inserindo os parâmetros suporte e consistência na Tabela 3, tem-se como resultado a Tabela 4.

A partir dos valores de suporte, que é calculado para cada regra, pode-se generalizar as informações, assunto que será abordado no próximo tópico.

Tabela 4: Sistema de Decisão com os Parâmetros φ e δ .

<i>U</i>	<i>C</i>					<i>D</i>		<i>Suporte</i>	<i>Consistência</i>
Criança	<i>Co</i>	<i>Ta</i>	<i>To</i>	<i>Te</i>	<i>Ma</i>	<i>At</i>	NRI	φ	δ
1	<i>Az</i>	<i>Gr</i>	<i>Du</i>	<i>In</i>	<i>Pl</i>	<i>Ng</i>	1	1	1
2	<i>Vm</i>	<i>Me</i>	<i>Mo</i>	<i>Li</i>	<i>Md</i>	<i>Nt</i>	1	1	1
3	<i>Am</i>	<i>Pq</i>	<i>Mc</i>	<i>As</i>	<i>Pe</i>	<i>Po</i>	1	1	1
4	<i>Az</i>	<i>Me</i>	<i>Mo</i>	<i>As</i>	<i>Pl</i>	<i>Ng</i>	1	1	1
5	<i>Am</i>	<i>Pq</i>	<i>Mc</i>	<i>In</i>	<i>Pl</i>	<i>Nt</i>	1	1	1
6	<i>Vd</i>	<i>Gr</i>	<i>Du</i>	<i>Li</i>	<i>Md</i>	<i>Po</i>	1	1	0,5
7	<i>Am</i>	<i>Pq</i>	<i>Du</i>	<i>In</i>	<i>Mt</i>	<i>Po</i>	1	1	1
8	<i>Am</i>	<i>Pq</i>	<i>Du</i>	<i>In</i>	<i>Pl</i>	<i>Po</i>	1	1	1
9	<i>Vd</i>	<i>Gr</i>	<i>Du</i>	<i>Li</i>	<i>Md</i>	<i>Nt</i>	1	1	0,5
10	<i>Vd</i>	<i>Me</i>	<i>Mo</i>	<i>Li</i>	<i>Pl</i>	<i>Nt</i>	1	1	1

2.1. Generalização da Informação

O propósito de generalização da informação é reduzir o número de classes de um sistema de decisão, de forma que a análise deste sistema fique mais fácil. Como o interesse é identificar, no banco de dados, classes que tenham suporte suficiente, é importante agrupar as regras com suporte insuficiente em uma única classe e, retirá-las do sistema de decisão a ser analisado, resultando em uma redução geral de número de classes neste sistema. Assim, para este procedimento, utiliza-se o valor do suporte. Neste caso, o φ é utilizado em \mathbb{R} para analisar a quantidade de objetos ou registros que classificam-se em cada valor nominal do conjunto do atributo de decisão $D = \{Nt, Po, Ng\}$.

No sistema de decisão \mathbb{R} da Tabela 4, existem dois registros com valor nominal do atributo de decisão $At = Ng$, quatro registro com $At = Po$ e quatro com $At = Nt$. Considerando a análise de suporte através do atributo de decisão, pode-se obter a Tabela 5.

Tabela 5: Valor de φ para Cada Valor Nominal do Atributo de Decisão (D).

<i>U</i>	<i>D</i>		<i>Suporte</i>
Criança	At	NRI	φ
(1, 4)	<i>Ng</i>	2	2
(2, 3, 5, 10)	<i>Nt</i>	4	4
(6, 7, 8, 9)	<i>Po</i>	4	4

Assumindo, para este sistema, o valor previamente definido por um especialista e, atendendo o propósito de generalização, precisa-se sinalizar os registros com suporte insuficiente para que estes sejam colocados em uma mesma classe. Este

procedimento tem por objetivo substituir o valor nominal do atributo de decisão, com $\varphi \leq 2$, do sistema de decisão R , por “-”. Através desta sinalização é possível particionar o sistema de decisão R em dois subsistemas: conjunto de classes com suporte superior ao suporte previamente definido, denotado por RR e, conjunto de classes com suporte igual ou inferior ao suporte definido, denotado por RI .

O subsistema de decisão RI contém todos os objetos com valores “-” para o atributo de decisão (Tabela 6), e o subsistema de decisão RR contém o restante dos objetos (Tabela 7). É bem provável que, a maioria das regras de classificação potencialmente úteis, sejam derivadas de RR . Esta é uma hipótese considerada razoável, pois, de outra maneira, considerando RI na geração das regras de classificação, a qualidade e quantidade das regras geradas tornam-se consideravelmente pequenas, devido à insuficiência de suporte para serem úteis. Este particionamento é viável para banco de dados que contenham grandes números de registros e muitos deles apareçam raramente [2]. O particionamento do sistema de decisão ajuda a economizar tempo computacional, pois resulta num sistema um pouco menor para obtenção de regras, quando esta é realizada sem a partição. Após o particionamento do banco de dados, inicia-se o processo de aplicação da técnica de Rough Sets para encontrar o reduto do sistema RR , assunto que será abordado a seguir.

Tabela 6: Subsistema de Decisão RI .

U	C					D		$Suporte$	$Consistência$
Criança	Co	Ta	To	Te	Ma	At	NRI	φ	δ
1	Az	Gr	Du	In	Pl	“-”	1	1	1
4	Az	Me	Mo	As	Pl	“-”	1	1	1

Tabela 7: Subsistema de Decisão RR .

U	C					D		$Suporte$	$Consistência$
Criança	Co	Ta	To	Te	Ma	At	NRI	φ	δ
2	Vm	Me	Mo	Li	Md	Nt	1	1	1
3	Am	Pq	Mc	As	Pe	Po	1	1	1
5	Am	Pq	Mc	In	Pl	Nt	1	1	1
6	Vd	Gr	Du	Li	Md	Po	1	1	0,5
7	Am	Pq	Du	In	Mt	Po	1	1	1
8	Am	Pq	Du	In	Pl	Po	1	1	1
9	Vd	Gr	Du	Li	Md	Nt	1	1	0,5
10	Vd	Me	Mo	Li	Pl	Nt	1	1	1

2.2. Redução do Sistema de Decisão

A redução do sistema de decisão é idêntico ao método de redução do sistema de informação demonstrado em [3], mas, agora, a redução é obtida a partir de RR ao invés de aplicar a todo o sistema R , uma vez que a maioria das regras de classificação potencialmente úteis pertence a este subsistema, o qual satisfaz o requisito de suporte definido pelo especialista ($\varphi > 2$).

Considerando $C = \{Co, Ta, To, Te, Ma\}$ e aplicando os conceitos de Rough Sets à Tabela 7, tem-se o conjunto dos atributos mínimos para a classificação do subsistema RR , formado por:

$$RED(C) = \{\{To, Ma\}\}.$$

Considerando o reduto $\{To, Ma\}$, tem-se então a redução do subsistema RR representada na Tabela 8. Ou seja, tem-se apenas os atributos condicionais necessários para manter as mesmas características do sistema original RR . A redução é simples, entretanto, a atenção deve ser redobrada, pois, as regras descobertas desta maneira não garantem total consistência com o sistema de decisão original, Tabela 3. Isso ocorre porque as regras podem satisfazer os atributos condicionais do subsistema de decisão RI mas não atendem os valores dos atributos de decisão. Ainda assim, as regras, após a redução do subsistema de decisão RR são válidas.

Observe a Tabela 9. Esta foi montada a partir do RI (Tabela 6), considerando, agora, o reduto encontrado para o subsistema RR , $\{To, Ma\}$. Nesta Tabela, as crianças 1 e 4 apresentam as mesmas características para os atributos condicionais que as crianças 8 e 10, respectivamente, da Tabela 8. Estas crianças diferem apenas no atributo de decisão. Sendo assim, as regras 8 e 10 só podem ser aceitas pelo sistema de decisão se mantiverem as condições especificadas pelo especialista (suporte e consistência).

Tabela 8: Redução do Subsistema de Decisão RR .

U	C		D		$Suporte$	$Consistência$
Criança	To	Ma	At	NRI	φ	δ
2	Mo	Md	Nt	1	1	1
3	Mc	Pe	Po	1	1	1
5	Mc	Pl	Nt	1	1	1
6	Du	Md	Po	1	1	0,5
7	Du	Mt	Po	1	1	1
8	Du	Pl	Po	1	1	1
9	Du	Md	Nt	1	1	0,5
10	Mo	Pl	Nt	1	1	1

Na próxima seção é apresentado um procedimento de como generalizar o conjunto de regras de classificação desta redução que satisfaça os requisitos de suporte e consistência.

Tabela 9: Redução do Subsistema de Decisão *RI*.

<i>U</i>	<i>C</i>		<i>D</i>		<i>Suporte</i>	<i>Consistência</i>
Criança	To	Ma	At	NRI	φ	δ
1	<i>Du</i>	<i>Pl</i>	“-”	1	1	1
4	<i>Mo</i>	<i>Pl</i>	“-”	1	1	1

2.3. Geração de Regras

Para melhor esclarecimento e simplificação, a geração de regras será realizada em duas partes:

- Geração de regras
- Redução de regras

Neste trabalho, optou-se por eliminar algumas condições desnecessárias contidas nas regras reduzidas, obtidas pela redução do subsistema de decisão *RR*. Obteve-se, assim, um conjunto de regras gerais minimizadas, ou seja, regras com o máximo de condições desnecessárias removidas. Isso foi feito suprimindo uma condição por vez para cada regra e verificando se a relação reduzida ainda permanecia consistente. Se a regra continuar classificando a mesma quantidade de objetos no subsistema de decisão *RR*, a condição é suprimida da regra.

A Tabela 10 mostra um conjunto de regras gerais minimizadas, denotado por *RR_{min}*, correspondente à redução obtida a partir de *RR* (Tabela 8), utilizando o princípio de eliminação de condições desnecessárias descritos anteriormente.

Tabela 10: Redução do Subsistema de Decisão *RR_{min}*.

<i>U</i>	<i>C</i>		<i>D</i>		<i>Suporte</i>	<i>Consistência</i>
Criança	To	Ma	At	NRI	φ	δ
(2, 10)	<i>Mo</i>	“-”	<i>Nt</i>	2	2	1
3	“-”	<i>Pe</i>	<i>Po</i>	1	1	1
5	<i>Mc</i>	<i>Pl</i>	<i>Nt</i>	1	1	1
6	<i>Du</i>	<i>Md</i>	<i>Po</i>	1	1	0,5
7	“-”	<i>Mt</i>	<i>Po</i>	1	1	1
8	<i>Du</i>	<i>Pl</i>	<i>Po</i>	1	1	1
9	<i>Du</i>	<i>Md</i>	<i>Nt</i>	1	1	0,5

O “-” na Tabela 10 indica que é uma condição não importante e, conseqüentemente, desnecessária para a regra (condição é suprimida). Realiza-se, então, após a definição de *RR_{min}*, uma redução de regras para eliminar as regras de classificação que não satisfaçam os requisitos de suporte e consistência definidos pelo especialista, denotado por *RR_{validas}*. A obtenção desta nova tabela segue os seguintes passos:

primeiramente, observa-se o suporte suficiente para o sistema, ou seja, só considera-se as regras que respeitam o suporte previamente definido. Posteriormente, para as regras admitidas no procedimento anterior, verifica-se a consistência, avaliando RR_{min} com o subsistema de decisão RI , pois este contém objetos cobrindo classes com suporte insuficiente. A consistência é facilmente medida por δ e pelo número de objetos de que satisfaz apenas os atributos condicionais da regra considerada. Levando-se em consideração os valores de suporte e consistência, definidos, neste momento, como sendo $\varphi \geq 1$ e $\delta \geq 0,5$, obtém-se da Tabela 10 o conjunto de regras válidas, conforme Tabela 11.

Tabela 11: $RR_{validas}$ para $\varphi \geq 1$ e $\delta \geq 0,5$.

<i>U</i>	<i>C</i>		<i>D</i>		<i>Suporte</i>	<i>Consistência</i>
Criança	To	Ma	At	NRI	φ	δ
R_1	<i>Mo</i>	"-"	<i>Nt</i>	2	2	0,667
R_2	"-"	<i>Pe</i>	<i>Po</i>	1	1	1
R_3	<i>Mc</i>	<i>Pl</i>	<i>Nt</i>	1	1	1
R_4	<i>Du</i>	<i>Md</i>	<i>Nt</i>	1	1	0,5
R_5	"-"	<i>Mt</i>	<i>Po</i>	1	1	1

A alteração dos valores de suporte e consistência permite encontrar regras de classificação mais generalizadas, porém, menos precisas. Os valores ótimos para estes parâmetros dependem do grau de confiabilidade necessário para o problema a ser analisado.

As principais características dos procedimentos aqui apresentados são:

- uso específico de dois parâmetros no processo de aprendizagem, suporte e consistência;
- o parâmetro suporte é usado para generalizar um determinado sistema de informação;
- a redução da informação é efetuada somente no subconjunto do sistema de informação, que tem suficiente suporte do banco de dados;
- as regras podem ter uma certa inconsistência com os dados, desde de que elas sejam suficientemente consistentes com o esperado.

3. Considerações Finais

Neste trabalho, apresentou-se um método para aprendizagem e obtenção de conhecimento em banco de dados, a partir da teoria de Rough Sets, que mostrou-se capaz de gerar regras de classificação não-determinística. O nível de consistência a ser tolerado pelas regras é especificado pelo usuário, e depende do grau de confiabilidade necessário a ser considerado no problema que será analisado. A utilização de regras não-determinísticas para a classificação de dados, é particularmente útil para lidar

com dados “ruidosos”, gerando regras mais gerais e confiáveis para uma análise de dados.

Abstract. The objective of this paper is to present a method based on Rough Sets, capable of generating rules of non-deterministic classification, that enables the user to specify the minimum of consistency that the classification rule will have to meet, and only generate rules that meet this requirement. That means that, in this method, rules are required to give enough support and to be consistent only the necessary for the data bank.

Referências

- [1] D.J.A. Cid, “Mineração de Dados com Técnicas de Rough Sets”. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ, 2000.
- [2] C.M.M.M. Patrício, “Detecção de fraude ou erro de medição em grandes consumidores de energia elétrica utilizando Rough Sets baseados em dados históricos e dados em tempo real”. Dissertação de Mestrado, Universidade Federal de Mato Grosso do Sul, Campo Grande, MS, 2005.
- [3] C.M.M.M. Patrício, J.O.P. Pinto, C.C. Souza, Rough sets - Técnica de redução de atributos e geração de regras para classificação de dados, em “XXVIII Congresso Nacional de Matemática Aplicada e Computacional”, São Paulo, SP, 2005.
- [4] Z. Pawlaz, Rough sets, in “International Journal of Computer and Information Sciences”, p. 341-356, 1982.
- [5] Z. Pawlaz, “Rough Sets: Theoretical Aspects of Reasoning about Data”, Kluwer Academic Publishers, 1991.
- [6] N. Shan, W. Ziarko, H.J. Hamilton, N. Cercone, Using rough sets as tool for knowledge discovery, in “International Conference on Knowledge Discovery and Data Mining”, p.263-268, 1995.
- [7] R. Slowinski, “Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory”, Kluwer Academic Publishers, 1992.