

# DESENVOLVIMENTO DE MÉTODOS PARA REDUÇÃO DE DIMENSIONALIDADE DO ESPAÇO DE CARACTERÍSTICAS PARA RECONHECIMENTO DE PADRÕES

MARCOS FUZARO JÚNIOR<sup>1</sup>, DENISE GULIATO<sup>2</sup>

## RESUMO

Nas últimas duas décadas as imagens médicas têm sido fundamentais para detecção, diagnóstico e tratamento dos mais diversos tipos de patologias. A tarefa de reconhecimento de padrões requer a conversão de padrões presentes nas imagens em características que descrevem as regiões de interesse de forma concisa. No entanto, em aplicações médicas, o número de características pode se tornar muito grande, sendo que muitas delas podem não ser essenciais para a solução do problema. Sendo assim, a redução de dimensionalidade tem papel importante no desempenho de métodos de classificação. Este trabalho apresenta dois métodos de redução do espaço de características, um baseado em seleção de características (*Rough Sets*) e outro baseado em transformação de características (*PCA*).

**PALAVRAS CHAVE:** redução de características, seleção de características, transformação de características.

## ABSTRACT

In the last two decades, medical imaging have been an important tool in detection, diagnosis and treatment of several types of pathologies. The task of pattern recognition requires the conversion of patterns in characteristics that describe images in a concise manner. However, in medical applications, the number of features can become very large, many of which may not be essential for the solution of the problem. Therefore, dimensionality reduction plays an important role in the performance of classification methods. This paper presents two methods to reduce the space of characteristics, the first based on selection of features using *Rough Sets* and the second is based on *Principal Component Analysis (PCA)*.

<sup>1</sup>Faculdade de Computação, Universidade Federal de Uberlândia, Avenida João Naves de Ávila 2121, Uberlândia-MG, CEP: 38400-902, marcosjunior@comp.ufu.br

<sup>2</sup>Faculdade de Computação, Universidade Federal de Uberlândia, Avenida João Naves de Ávila 2121, Uberlândia-MG, CEP: 38400-902, guliato@ufu.br

**KEYWORDS:** reduction of features, feature selection, processing characteristics.

## 1. INTRODUÇÃO

O crescimento no número de modalidades de exames por imagens, e suas aplicações práticas, tem sido acompanhada por um aumento natural no escopo e na complexidade dos problemas associados aos métodos de aquisição, requerendo técnicas computacionais cada vez mais avançadas para sua solução.

O grande potencial no uso de sistemas computacionais aplicados a imagens médicas está associado ao uso de técnicas de processamento digital de imagens e de visão computacional para análise quantitativa e objetiva das imagens, no uso de Sistemas de Gerenciamento de Banco de Dados (SGBD) para armazenar grandes volumes de dados de natureza complexa e para recuperar tais informações de maneira eficiente, e no desenvolvimento de métodos de reconhecimento de padrões.

A tarefa de reconhecimento de padrão requer a conversão de padrões presentes nas imagens em características que descrevem as regiões de interesse de forma concisa. No entanto, em aplicações médicas, o número de características pode se tornar muito grande, com características não relevantes ou redundantes que podem não ser essenciais para a solução do problema. Sendo assim, a redução de dimensionalidade tem papel importante no desempenho de métodos de classificação, pois influencia diretamente no custo e na precisão do classificador.

Neste trabalho são apresentados dois métodos de redução da dimensionalidade do espaço de características [1]: um baseado em seleção de características utilizando *Rough Sets* [2], [3], [4] e outro baseado em transformação de características usando análise das componentes principais (*PCA*) [7], [8].

O trabalho está estruturado da seguinte forma: a Seção 2 apresenta o método baseado em seleção de características (*Rough Sets*); a Seção 3 apresenta o método baseado em transformação de características (*PCA*); a Seção 4 apresenta materiais e métodos; a Seção 5 apresenta uma comparação entre os dois métodos e a Seção 6 apresenta a conclusão do trabalho.

## 2. CONCEITOS BÁSICOS SOBRE A TEORIA DOS ROUGH SETS

A técnica de redução de dimensionalidade do espaço de características baseada em seleção de características é um processo que consiste em obter um subconjunto de características a partir de um conjunto de características originais baseado em determinado critério, de maneira que o subconjunto expresse as propriedades do conjunto original.

A teoria dos Rough Sets, introduzida por Zdzisław Pawlak em 1982, é um método baseado em seleção de características que permite eliminar características irrelevantes pela formação de redutos, que são subconjuntos que mantêm as propriedades da representação do conjunto original.

Neste trabalho, serão apresentados conceitos e definições seguidos de exemplos para facilitar o entendimento. As subseções 2.1, 2.2, 2.3 e 2.4 mostram conceitos e definições utilizados na teoria dos Rough Sets. A subseção 2.5 apresenta os passos necessários para obter os redutos.

### 2.1 Sistema de Informação

Os dados são representados por um sistema de informação composto por um conjunto de objetos e um conjunto de atributos. Todos os objetos possuem os mesmos atributos, diferindo somente em seus valores nominais. Portanto, esse sistema de informação pode ser definido como  $A = (U, C)$ , sendo  $U$  o conjunto de objetos e  $C$  o conjunto de atributos. A Tabela 1 [3] representa um exemplo de um sistema de informação, em que os dados são referentes a crianças e algumas características de brinquedos.

Tabela 1- Sistema de Informação [3]

U	Atributos Condicionais				
	Criança	Cor	Tamanho	Tato	Textura
1	Azul	Grande	Duro	Indefinido	Plástico
2	Vermelho	Médio	Moderado	Liso	Madeira
3	Amarelo	Pequeno	Macio	Áspero	Pelúcia
4	Azul	Médio	Moderado	Áspero	Plástico
5	Amarelo	Pequeno	Macio	Indefinido	Plástico
6	Verde	Grande	Duro	Liso	Madeira
7	Amarelo	Pequeno	Duro	Indefinido	Metal
8	Amarelo	Pequeno	Duro	Indefinido	Plástico
9	Verde	Grande	Duro	Liso	Madeira
10	Verde	Médio	Moderado	Liso	Plástico

Uma outra maneira de representar os dados em um sistema de informação é incluir um atributo de decisão que informa a decisão a ser tomada. Pode ser representado por  $R = (U, C \cup D)$ , onde  $D$  é o atributo de decisão. A Tabela 2 representa um exemplo de um sistema de informação com um atributo de decisão, em que esse atributo é referente ao comportamento das crianças.

Tabela 2 – Sistema de informação munido de atributo de decisão [3]

U	Atributos Condicionais					Atributo de Decisão
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 o sistema de informação ilustrado pela tabela 2, temos:

- conjunto de objetos:  $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$
- conjunto de atributos condicionais:  $C = \{\text{Cor, Tamanho, Tato, Textura, Material}\}$
- atributo de decisão:  $D = \{\text{Atitude}\}$
- classes:  $X_1 = \{\text{Negativa}\}$     $X_2 = \{\text{Neutra}\}$     $X_3 = \{\text{Positiva}\}$

## 2.2 Indiscernibilidade

Considerando o sistema de informação  $R$ , para todo subconjunto de atributos  $B \subseteq C$ , uma relação de equivalência, chamada relação de indiscernibilidade ( $IND$ ), pode ser definida como:

$$IND_R = \{(x, y) \in U^2 \mid \forall c \in B, c(x) = c(y)\} \quad (1)$$

Os objetos  $x$  e  $y$  são indiscerníveis entre si, ou seja, todos os valores de seus atributos são iguais com respeito ao subconjunto  $B$ .

O conjunto de todas as classes de equivalência determinadas por  $IND_R(B)$  é representado por  $U/IND_R(B)$ . Para a base de dados da Tabela 2, temos como exemplos:

- $IND_R(\{\text{Tamanho, Tato, Textura}\}) = \{6, 9\}$
- $U/IND_R(\{\text{Tamanho, Tato}\}) = \{\{1, 6, 9\}, \{2, 4, 10\}, \{3, 5\}, \{7, 8\}\}$

### 2.3 Aproximação de Conjuntos

Analisando a Tabela 2, pode-se dizer que qualquer criança com as características iguais as das crianças 1 e 4 terão atitude *Negativa*, assim como crianças com características iguais as crianças 2, 5 ou 10 terão atitude *Neutra* e qualquer criança com características iguais as das crianças 3, 7 ou 8 terão atitude *Positiva*. A respeito das crianças 6 e 9, nada se pode afirmar, pois possuem os mesmos atributos condicionais e atributos de decisão diferentes.

Seja o sistema de informação  $A = (U, C)$ ,  $B \subseteq C$  e  $X \subseteq U$ , onde  $X$  é o conjunto de elementos de uma classe. Defini-se Aproximação Superior (*Upper*) de  $X_i$  em relação a  $B$ , denotado por  $\overline{B}(X_i)$ , como:

$$\overline{B}(X) = \{x \in U \mid U/IND_A(B) \cap X \neq \emptyset\}$$

e Aproximação Inferior (*Lower*) de  $X$  em relação a  $B$ , denotado por  $\underline{B}(X_i)$ , como:

$$\underline{B}(X) = \{x \in U \mid U/IND_A(B) \subseteq X\}$$

Utilizando o conjunto de atributos  $B$ , os objetos pertencentes a Aproximação Superior de  $X_i$  são classificados como possíveis membros de  $X_i$ , e os objetos da Aproximação Inferior de  $X_i$  são objetos que com certeza pertencem a  $X_i$ . A partir disso, obtém-se uma Região de Fronteira  $RF(X_i)$ , denotada por:

$$RF(X) = \overline{B}(X) - \underline{B}(X)$$

e ainda define-se Fora da Região de  $X_i$  o conjunto  $U - \overline{B}(X_i)$ , que consiste de objetos que não pertencem a  $X_i$ , considerando o mesmo conjunto  $B$ . A Figura 1 [2] ilustra esses conceitos.

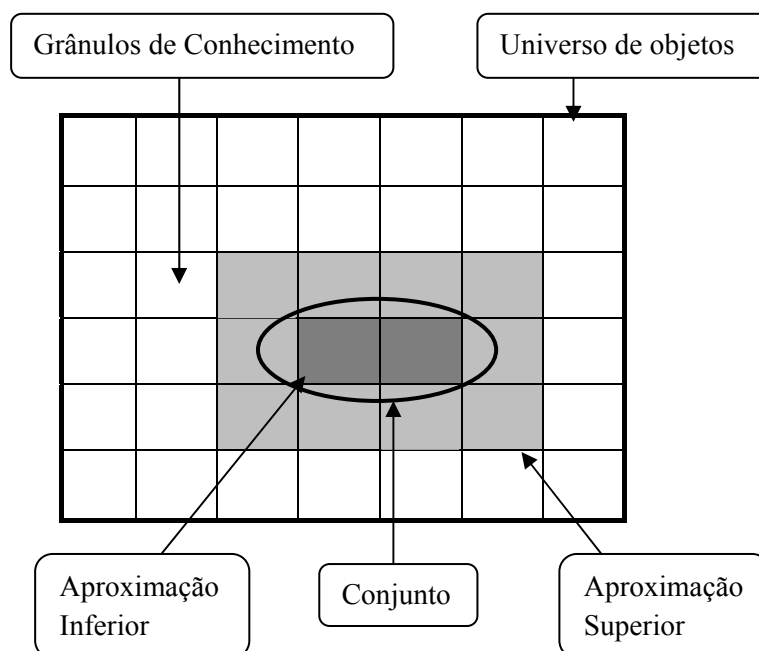


Figura 1 – Aproximação de conjuntos [2]

Um conjunto  $X_i$  é definido como *rough* (impreciso) se sua Região de Fronteira é diferente do conjunto vazio, e é definido como *crisp* (preciso) se sua Região de Fronteira for vazia.

Vamos tomar um exemplo para ilustrar os conceitos acima descritos [3]. Seja  $B = \{\text{Cor, Tamanho, Tato, Textura, Material}\}$ ,  $U/IND_A(B) = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6, 9\}, \{7\}, \{8\}, \{10\}\}$  e  $X_1 = \{3, 6, 7, 8\}$ ,  $X_1$  é o conjunto formado por crianças que possuem atitude Positiva. Tem-se os seguintes conjuntos, ilustrados na Figura 2:

- Aproximação Superior:  $\bar{B}(X_1) = \{\{3\}, \{6, 9\}, \{7\}, \{8\}\}$
- Aproximação Inferior:  $\underline{B}(X_1) = \{\{3\}, \{7\}, \{8\}\}$
- Região de Fronteira:  $RF(X_1) = \{\{6, 9\}\}$
- Fora da Região  $U - \bar{B}(X_1) = \{\{1\}, \{2\}, \{4\}, \{5\}, \{10\}\}$

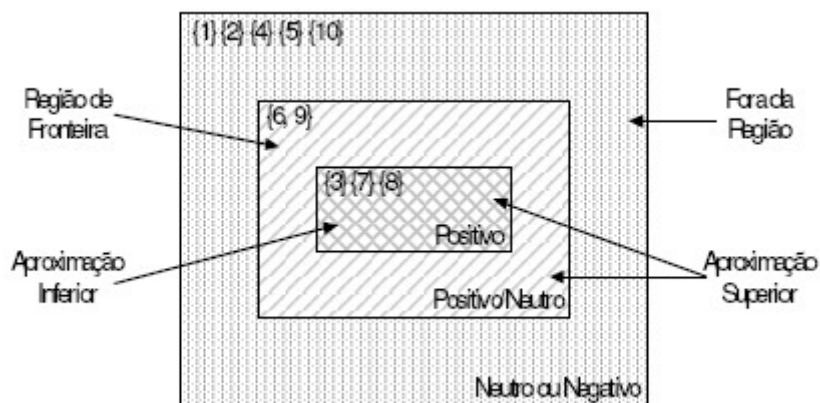


Figura 2 – Representação em conjuntos (exemplo) [3]

- Fora da Região
- Região de Fronteira
- Aproximação Inferior
- Aproximação Superior  $\cup$

## 2.4 Qualidade das Aproximações

As qualidades das aproximações definidas anteriormente podem ser caracterizadas por valores numéricos. Seja  $\alpha_B(X)$  o coeficiente para medir essas qualidades, sendo  $X$  o conjunto de objetos com respeito a  $B$ . Esse coeficiente pode ser calculado de três maneiras:

- a) Coeficiente de Imprecisão  $\alpha_B(X)$ , que é dado pelo quociente entre a cardinalidade das aproximações Inferior e Superior. Denota-se por:

$$\alpha_B(X) = \frac{|B(X)|}{|\overline{B}(X)|}$$

Se  $\alpha_B(X) = 1$ ,  $X$  é preciso em relação ao conjunto de atributos  $B$ , e se  $\alpha_B(X) < 1$ ,  $X$  é impreciso em relação ao conjunto de atributos  $B$ .

- b) Coeficiente da Qualidade da Aproximação Superior  $\alpha_B(\overline{B}(X))$ , que pode ser interpretado como a porcentagem de objetos que possivelmente pertencem a  $X$ , dado por:

$$\alpha_B(\overline{B}(X)) = \frac{|\overline{B}(X)|}{|U|}$$

Sendo  $|U|$  a cardinalidade do conjunto de objetos do Sistema de Informação e  $U \neq \{ \}$ .

- c) Coeficiente da Qualidade da Aproximação Inferior  $\alpha_B(\underline{B}(X))$ , que pode ser interpretado como a porcentagem de objetos que com certeza pertencem a  $X$ , dado por:

$$\alpha_B(\underline{B}(X)) = \frac{|\underline{B}(X)|}{|U|}$$

Como exemplo vamos tomar  $X_1 = \{3, 6, 7, 8\}$ , tem-se:

- $\alpha_B(X_1) = \frac{|\{3, 7, 8\}|}{|\{3, 6, 7, 8, 9\}|} = \frac{3}{5} = 0.6$ , ou seja, 60% de  $X_1$  é preciso em relação a  $B$ .
- $\alpha_B(\overline{B}(X_1)) = \frac{|\{3, 6, 7, 8, 9\}|}{|\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}|} = \frac{5}{10} = 0.5$ , ou seja, 50% de  $U$  possivelmente pertence a  $X_1$ .
- $\alpha_B(\underline{B}(X_1)) = \frac{|\{3, 7, 8\}|}{|\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}|} = \frac{3}{10} = 0.3$ , ou seja, 30% de  $U$  certamente pertence a  $X_1$ .

## 2.5 Redução do Sistema de Informação

Um sistema de informação não deve possuir características irrelevantes, pois isso aumenta a complexidade computacional da extração de regras para classificação. Por isso, quando possível, deve-se reduzir o sistema de informação. Na teoria dos *Rough Sets*, essa redução é realizada pela função de discernibilidade a partir da matriz de discernibilidade.



### 2.5.1 Matriz de Discernibilidade

A Matriz de Discernibilidade de um sistema de informação A, denotada por  $M_D(B)$ , é uma matriz simétrica  $n \times n$  com:

$$m_D(i, j) = \{b \in B \mid b(Cl(i)) \neq b(Cl(j))\}$$

para  $i, j = 1, 2, 3, \dots, n$ , sendo  $1 \leq i, j \leq n$

$$n = |U/IND_A(B)|$$

e  $Cl(n)$  é a classe de equivalência  $n$

Considere o sistema de informação A da Tabela 1 e seja  $B = \{\text{Cor, Tamanho, Tato, Textura, Material}\}$ . Tem-se  $U/IND_A(B) = \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6, 9\}, \{7\}, \{8\}, \{10\}\}$ . Portanto, cada elemento dessa matriz é o conjunto de atributos condicionais de B que diferenciam as classes de equivalência. Considerando  $\text{Cor} = \text{Co}$ ,  $\text{Tamanho} = \text{Ta}$ ,  $\text{Tato} = \text{To}$ ,  $\text{Textura} = \text{Te}$  e  $\text{Material} = \text{Ma}$ , tem-se na Tabela 3 sua representação.

Tabela 3 – Matriz de Discernibilidade [3]

	CI1	CI2	CI3	CI4	CI5	CI6	CI7	CI8	CI9
CI1	$\emptyset$								
CI2	Co,Ta,To, Te,Ma	$\emptyset$							
CI3	Co,Ta,To, Te,Ma	Co,Ta,To, Te,Ma	$\emptyset$						
CI4	Ta,To,Te	Co,Te,Ma	Co,Ta,To, Ma	$\emptyset$					
CI5	Co,Ta,To	Co,Ta,To, Te,Ma	Te,Ma	Co,Ta,To, Te	$\emptyset$				
CI6	Co,Te,Ma	Co,Ta,To	Co,Ta,To, Te,Ma	Co,Ta,To, Te, Ma	Co,Ta,To, Te, Ma	$\emptyset$			
CI7	Co,Ta,Ma	Co,Ta,To, Te,Ma	To,Te,Ma	Co,Ta,To, Te, Ma	To,Ma	Co,Ta,Te, Ma	$\emptyset$		
CI8	Co,Ta	Co,Ta,To, Te,Ma	To,Te,Ma	Co,Ta,To, Te	To	Co,Ta,Te, Ma	Ma	$\emptyset$	
CI9	Co,Ta,To, Te	Co,Ma	Co,Ta,To, Te,Ma	Co,Te	Co,Ta,To, Te	Ta,To,Ma	Co,Ta,To, Te,Ma	Co,Ta,To, Te	$\emptyset$

### 2.5.2 Função de Discernibilidade

A Função de discernibilidade  $F_A(B)$  determina o número mínimo de atributos para diferenciar qualquer classe de equivalência das demais. É obtida da seguinte maneira: para os atributos contidos dentro de cada célula da matriz de discernibilidade, aplica-se o operador “or” ou “ $\vee$ ”, e entre as células dessa matriz, aplica-se o operador “and” ou “ $\wedge$ ”, o que resulta

em uma expressão booleana. Em seguida, aplicando teoremas e propriedades da Álgebra booleana, simplifica-se a expressão.

Vamos ilustrar esse processo através de um exemplo. A  $F_A(B)$  da Tabela 3 é dada por:

$$F_A(B) = (Co \vee Ta \vee To \vee Te \vee Ma) \wedge (Co \vee Ta \vee To \vee Te \vee Ma) \wedge (Ta \vee To \vee Te) \wedge (Co \vee Ta \vee To) \wedge (Co \vee Te \vee Ma) \wedge (Co \vee Ta \vee Ma) \wedge (Co \vee Ta) \wedge (Co \vee Ta \vee To \vee Te) \wedge (Co \vee Ta \vee To \vee Te \vee Ma) \wedge (Co \vee Te \vee Ma) \wedge (Co \vee Ta \vee To \vee Te \vee Ma) \wedge (Co \vee Ta \vee To) \wedge (Co \vee Ta \vee To \vee Te \vee Ma) \wedge (Co \vee Ta \vee To \vee Te \vee Ma) \wedge (Co \vee Ma) \wedge (Co \vee Ta \vee To \vee Ma) \wedge (Te \vee Ma) \wedge (Co \vee Ta \vee To \vee Te \vee Ma) \wedge (To \vee Te \vee Ma) \wedge (To \vee Te \vee Ma) \wedge (Co \vee Ta \vee To \vee Te \vee Ma) \wedge (Co \vee Ta \vee To \vee Te) \wedge (Co \vee Ta \vee To \vee Te \vee Ma) \wedge (Co \vee Ta \vee To \vee Te \vee Ma) \wedge (Co \vee Ta \vee To \vee Te) \wedge (Co \vee Te) \wedge (Co \vee Ta \vee To \vee Te \vee Ma) \wedge (To \vee Ma) \wedge (To) \wedge (Co \vee Ta \vee To \vee Te) \wedge (Co \vee Ta \vee Te \vee Ma) \wedge (Co \vee Ta \vee Te \vee Ma) \wedge (Ta \vee To \vee Ma) \wedge (Ma) \wedge (Co \vee Ta \vee To \vee Te \vee Ma) \wedge (Co \vee Ta \vee To \vee Te)$$

Simplificando essa expressão, tem-se:

$$F_A(B) = ((Co \vee Ta \vee Te) \wedge To \wedge Ma)$$

que pode ser escrita na forma:

$$F_A(B) = ((Co \wedge To \wedge Ma) \vee (Ta \wedge To \wedge Te \wedge Ma))$$

### 2.5.3 Reduto

Reduto é um conjunto de atributos mínimos que mantém as propriedades de um sistema de informação. Um reduto de B sobre um sistema de informação A é um conjunto de atributos  $B^*$ ,  $B^* \subseteq B$ , sendo todos os atributos  $c \in (B - B^*)$  dispensáveis. Com isso,  $U/IND_A(B) = U/IND_A(B^*)$ .

Considerando  $B = \{\text{Cor, Tamanho, Tato, Textura, Material}\}$  e a função de discernibilidade obtida anteriormente, o conjunto de redutos dessa função é:

$$RED_A(B) = \{\{\text{Cor, Tato, Material}\}, \{\text{Tamanho, Tato, Textura, Material}\}\}$$

O sistema de informação original, ilustrado na Tabela 1, pode ser representado por um dos redutos. Utilizando o reduto  $B^* = \{\text{Cor, Tato, Material}\}$ , temos a redução do sistema de informação, indicado na tabela 4, pois ao invés de termos a tabela original com 5 atributos temos uma representação equivalente com 3 atributos.

Tabela 4 – Redução do sistema de informação [3]

U	Atributos Condicionais		
	Cor	Tato	Material
1	<i>Azul</i>	<i>Duro</i>	<i>Plástico</i>
2	<i>Vermelho</i>	<i>Moderado</i>	<i>Madeira</i>
3	<i>Amarelo</i>	<i>Macio</i>	<i>Pelúcia</i>
4	<i>Azul</i>	<i>Moderado</i>	<i>Plástico</i>
5	<i>Amarelo</i>	<i>Macio</i>	<i>Plástico</i>
6	<i>Verde</i>	<i>Duro</i>	<i>Madeira</i>
7	<i>Amarelo</i>	<i>Duro</i>	<i>Metal</i>
8	<i>Amarelo</i>	<i>Duro</i>	<i>Plástico</i>
9	<i>Verde</i>	<i>Duro</i>	<i>Madeira</i>
10	<i>Verde</i>	<i>Moderado</i>	<i>Plástico</i>

### 3. PRINCIPAL COMPONENT ANALYSIS (PCA)

O método baseado em transformação de características (PCA) [7], [8], descrito pela primeira vez por Karl Pearson em 1901, é um processo que representa em uma nova projeção para os dados, reduzindo a dimensionalidade do espaço. A partir de características originais, que estão correlacionadas, obtém-se um novo conjunto de características descorrelacionadas.

Antes de descrever o método PCA são expostos conceitos matemáticos que são utilizados em seus passos. Esse conhecimento apriori é fundamental para um bom entendimento do PCA.

A subseção 3.1 apresenta os conceitos matemáticos. Na subseção 3.2 são apresentados os passos realizados pelo PCA para a redução de dimensionalidade.

#### 3.1 Conceitos Matemáticos

O objetivo ao citar esses conceitos não é fornecer um entendimento completo do mecanismo matemático de cada técnica, mas dizer o motivo da técnica ser utilizada e quais os resultados gerados e sua influência nos dados.

### 3.1.1 Desvio Padrão

O desvio padrão é a média das distâncias entre cada ponto e a média central. É uma medida estatística que indica a dispersão dos dados, calculado pela seguinte fórmula:

$$s = \sqrt{\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)}}$$

onde:

$X_i$  = valor  $i$  da amostra

$n$  = número de valores da amostra

$\bar{X}$  = média dos valores

### 3.1.2 Variância

É uma outra medida de dispersão obtida pelo quadrado do desvio padrão:

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{(n-1)} = \frac{\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})}{(n-1)}$$

### 3.1.3 Covariância

As medidas citadas anteriormente, variância e desvio padrão, são utilizadas quando os dados possuem 1 dimensão. Ao se tratar de mais dimensões, a medida que deve ser utilizada é a covariância, pois ela fornece o comportamento de uma dimensão em relação a outra. Por exemplo, em situações de 3 dimensões ( $x$ ,  $y$ ,  $z$ ), pode-se medir a covariância entre as dimensões  $x$  e  $y$ , as dimensões  $x$  e  $z$ , e as dimensões  $y$  e  $z$ . Assim, a covariância sempre é calculada entre duas dimensões.

A fórmula da covariância é similar a da variância:

$$\text{cov}(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

Ao calcular a covariância, o mais importante não é exatamente seu valor, e sim o seu sinal (positiva ou negativa). Se a covariância é positiva, indica que ambas dimensões aumentam juntas. Se a covariância é negativa, indica que enquanto uma dimensão aumenta, a

outra diminui. E se a covariância vale zero, então as dimensões não estão relacionadas entre si.

Propriedades da covariância:

- (i)  $cov(x,y) = cov(y,x)$
- (ii)  $cov(x,x) = var(x)$ .

### 3.1.4 Matriz de Covariância

Ao se tratar de  $n$  dimensões, uma maneira útil de obter todos os valores de covariância entre todas as diferentes dimensões é calculá-los e colocá-los em uma matriz. A definição para matriz de covariância para um conjunto de dados de  $n$  dimensões é:

$$C^{n \times n} = (c_{ij} / c_{ij} = cov(Dim_i, Dim_j))$$

onde:

$C^{n \times n}$  é a matriz com  $n$  linhas e  $n$  colunas, e  $Dim_x$  é a  $x$ -ésima dimensão.

Como exemplo, temos uma matriz de covariância para um conjunto de dados de 3 dimensões ( $x$ ,  $y$  e  $z$ ). Então, a matriz possui 3 linhas e 3 colunas, e seus valores são:

$$C = \begin{pmatrix} cov(x, x) & cov(x, y) & cov(x, z) \\ cov(y, x) & cov(y, y) & cov(y, z) \\ cov(z, x) & cov(z, y) & cov(z, z) \end{pmatrix}$$

### 3.1.5 Autovetores

Autovetores são vetores que sob a ação de um operador linear resultam num vetor de mesma direção, ou seja, representam uma direção que é preservada por uma transformação linear.

Os autovetores são encontrados somente em matrizes quadradas, e se uma matriz  $n \times n$  tem autovetores, então ela tem  $n$  autovetores. Por exemplo, uma matriz  $4 \times 4$  tem 4 autovetores. Uma propriedade que deve ser destacada é que todos os autovetores de uma matriz são perpendiculares entre si.

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

Figura 3 – Exemplo de um autovetor [7]

Na Figura 3 o vetor  $(3 \ 2)$  é um autovetor, pois o resultado da multiplicação é um vetor múltiplo de  $(3 \ 2)$ , que possui a mesma direção.

### 3.1.6 Autovalores

Autovalores e autovetores estão associados entre si e sempre aparecem em pares. Na Figura 3, o 4 é o autovalor associado ao autovetor  $(3 \ 2)$ .

## 3.2 Principal Component Analysis (PCA)

O método PCA permite encontrar padrões em uma base de dados, principalmente quando os dados possuem grandes dimensões e não é possível representá-los graficamente. Depois de encontrados esses padrões, é possível reduzir o número de dimensões sem perder muita informação. São encontrados autovetores e autovalores que norteiam o comportamento dos dados. A partir disso são selecionadas as componentes principais, que são os autovetores associados aos maiores autovalores.

Os passos que descrevem o método vão ser descritos e ilustrados com exemplos, facilitando o entendimento.

### 3.2.1 Passo 1: Obter os dados

Para descrever o processo realizado pelo PCA será utilizado uma base de dados bidimensional  $(x, y)$ , representada na Figura 4 [7].

x	y	classe
2.5	2.4	1
0.5	0.7	2
2.2	2.9	1
1.9	2.2	1
3.1	3.0	1
2.3	2.7	1
2.0	1.6	1
1.0	1.1	2
1.5	1.6	2
1.1	0.9	2

Figura 4 – Dados originais [7]

### 3.2.2 Passo 2: Subtrair a média

Depois de obter a base de dados é necessário ajustá-la subtraindo a média respectiva a cada dimensão. No exemplo, deve-se obter a média da dimensão  $x$  e subtraí-la de todos os valores dessa dimensão, e fazer o mesmo para a dimensão  $y$ . O resultado irá produzir uma nova base cuja média vale zero. A Figura 5 ilustra os dados ajustados.

x	y	classe
0.69	0.49	1
-1.31	-1.21	2
0.39	0.99	1
0.09	0.29	1
1.29	1.09	1
0.49	0.79	1
0.19	-0.31	1
-0.81	-0.81	2
-0.31	-0.31	2
-0.71	-1.01	2

Figura 5 – Dados com a média subtraída [7]

### 3.2.3 Passo 3: Calcular a matriz de covariância

Tem-se uma base bidimensional, então a matriz de covariância será  $2 \times 2$ . Seguindo os conceitos citados sobre matriz de covariância, tem-se o resultado abaixo:

$$\text{cov} = \begin{pmatrix} 0.616555556 & 0.615444444 \\ 0.615444444 & 0.716555556 \end{pmatrix}$$

Note que como os valores que não fazem parte da diagonal principal são positivos, indicando que ambas dimensões crescem juntas.

### 3.2.4 Passo 4: Calcular os autovetores e autovalores

Essa etapa é muito importante, pois irá fornecer detalhes sobre o comportamento dos dados. Calculando os autovetores e autovalores, segundo os conceitos apresentados, tem-se:

$$\text{autovalor}_1 = (0.0490833989)$$

$$\text{autovalor}_2 = (1.28402771)$$

$$\text{autovetor}_1 = \begin{pmatrix} -0.735178656 \\ 0.677873399 \end{pmatrix}$$

$$\text{autovetor}_2 = \begin{pmatrix} -0.677873399 \\ -0.735178656 \end{pmatrix}$$

Observando a Figura 6, tem-se que as dimensões crescem juntas, como foi mencionado anteriormente. Os autovetores estão representados pelas diagonais pontilhadas, e são perpendiculares entre si. Veja como os autovetores passam pelos pontos mostrando como os dados estão relacionados entre si. Tomando como base o autovetor 1 e a projeção dos pontos neste vetor, vemos que os dados transformados tem uma boa representação. Utilizando o autovetor 2, ao projetar os dados sobre esse autovetor irá ocorrer sobreposição dos dados, causando problemas na representação. Portanto, o método de análise de componentes principais depende fortemente da distribuição dos dados.

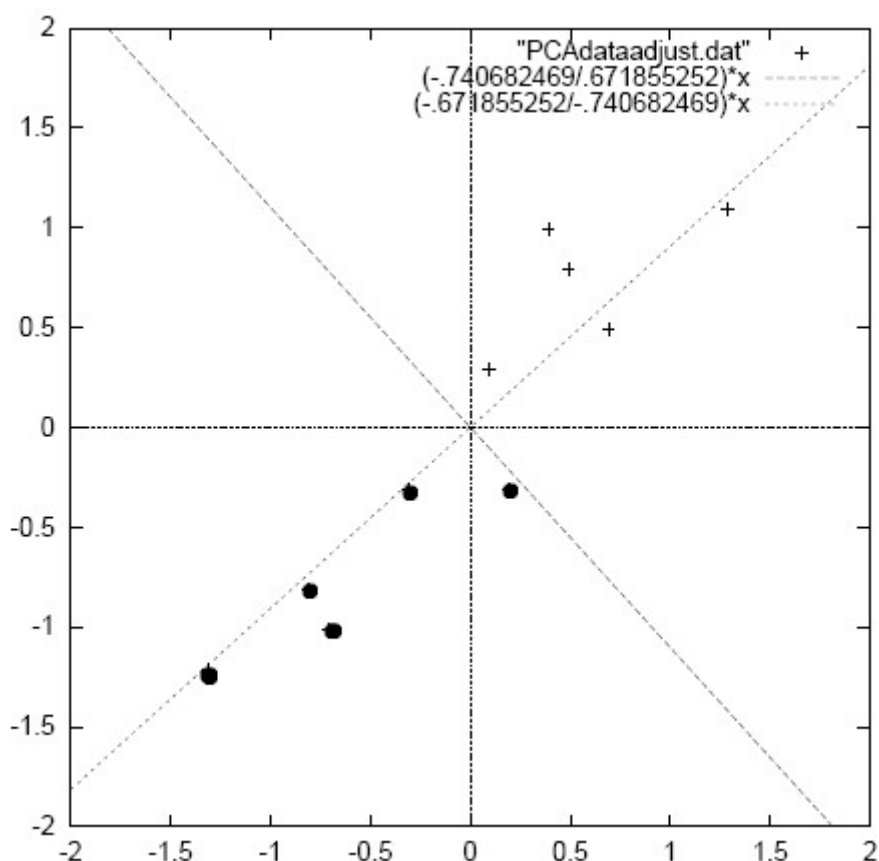


Figura 6 – Gráfico de dados normalizados (média subtraída), e com os autovetores [7]



### 3.2.5 Passo 5: Escolher as componentes e formar o vetor de características

É nessa etapa que emerge o conceito de componentes principais. O autovetor associado ao maior autovalor é a componente principal mais relevante. Quanto maior o autovalor associado, maior é a importância do autovetor (componente). Assim, a redução de dimensionalidade ocorre ao retirar as componentes menos significantes. No exemplo, o autovetor mais importante é o que passa no meio dos pontos.

$$\begin{pmatrix} -0.677873399 \\ -0.735178656 \end{pmatrix}$$

Ao escolher as componentes, deve-se construir o vetor de características (*FeatureVector*), que é um vetor de vetores.

$$FeatureVector = (av_1 \ av_2 \ av_3 \ \dots \ av_n)$$

onde  $av_i$  é o autovetor  $i$

No nosso exemplo, o vetor de características fica da seguinte maneira:

$$FeatureVector = \begin{pmatrix} -0.677873399 \\ -0.735178656 \end{pmatrix}$$

### 3.2.6 Formando o novo conjunto de dados

O novo conjunto de dados é dado multiplicando a transposta do vetor de características pela transposta dos dados ajustados (com a média subtraída).

$$FinalData = FeatureVector^T \times DataAdjust^T$$

No exemplo, a base original tem duas dimensões. Escolhendo um autovetor, o de maior autovalor, obtemos um novo conjunto de dados com 1 dimensão.

x	classe
-0.827970186	1
1.77758033	2
-0.992197494	1
-0.274210416	1
-1.67580142	1
-0.912949103	1
0.0991094375	1
1.14457216	2
0.438046137	2
1.22382056	2

#### 4. MATERIAIS E MÉTODOS

Através do software RSES (Rough Set Exploration System) [9], utilizando o método de seleção de características (*rough sets*) [2], [3], [4], e o software R-Project [10], utilizando o método de transformação de características denominado PCA [7], [8], foram utilizados para executar testes nas seguintes bases de dados públicas: *Iris database* com 4 características, *Wine database* com 13 características, *Wisconsin Diagnostic Breast Cancer database* com 10 características, *Wisconsin Diagnostic Breast Cancer database* com 30 características e *Wisconsin Prognostic Breast Cancer* com 9 características.

Em seguida, as bases de dados pré-processadas foram submetidas ao Classificador JADE baseado em regras fuzzy desenvolvido no nosso laboratório [5], [6] e aos classificadores: BayesNet, Decision Table, J48, LibSVM e Multilayer Perceptron, disponíveis no ambiente Weka [11].

Ao final, foram montadas tabelas que permitem avaliar, de maneira clara, a eficiência entre os métodos.

#### 5. DISCUSSÃO DE RESULTADOS

Fazendo uma análise nos resultados obtidos nos testes, o método baseado em seleção de características utilizando *rough sets* [2], [3], [4] apresentou redução de dimensionalidade do espaço de características em algumas bases de dados utilizadas. A *Wine database* com 13

características foi reduzida para 7 características e a *Wisconsin Diagnostic Breast Cancer database* de 30 características foi reduzida para 16 características. As demais bases não sofreram redução de dimensionalidade. O método baseado em transformação de características (PCA) [7], [8] foi aplicado a todas as bases, gerando novas características. Foram escolhidas metade das componentes principais, as mais relevantes, para a realização dos testes.

Em seguida, as bases de dados pré-processadas foram submetidas aos seguintes classificadores: Classificador JADE baseado [5], [6], BayesNet, Decision Table, J48, LibSVM e Multilayer Perceptron. Os resultados obtidos se encontram nas tabelas abaixo.

Tabela 5 – *Iris database*

<b>IRIS</b>	base original (4 caract)	base PCA (2 comp)	base Redutos (4 caract)
JADE	100%	100%	100%
BayesNet	92,66%	90,66%	92,66%
Decision Table	92,66%	91,36%	92,66%
J48	96,03%	94,16%	96,03%
LibSVM	96,03%	91,56%	96,03%
Multilayer Perceptron	97,33%	96,80%	97,33%

Tabela 6 – *Wine database*

<b>WINE</b>	base original (13 caract)	base PCA (7 comp)	base Redutos (7 caract)
JADE	99,04%	89,82%	76,16%
BayesNet	98,76%	91,00%	93,50%
Decision Table	92,90%	86,06%	87,80%
J48	94,33%	86,60%	89,73%
LibSVM	97,76%	94,60%	97,60%
Multilayer Perceptron	97,03%	95,36%	96,00%

Tabela 7 - *Wisconsin Prognostic Breast Cancer Database*

<b>WPBC</b>	base original (9 caract)	base PCA (5 comp)	base Redutos (9 caract)
JADE	93,09%	95,90%	93,09%
BayesNet	96,30%	95,85%	96,30%
Decision Table	92,65%	95,80%	92,65%
J48	93,85%	96,25%	93,85%
LibSVM	96,05%	96,40%	96,05%
Multilayer Perceptron	95,80%	95,80%	95,80%

Tabela 8 - *Wisconsin Diagnostic Breast Cancer Database*

WDBC	base original (10 caract)	base PCA (5 comp)	base Redutos (10 caract)
JADE	97,45%	19,43%	97,45%
BayesNet	92,40%	91,80%	92,40%
Decision Table	93,15%	90,65%	93,15%
J48	92,65%	88,90%	92,65%
LibSVM	93,80%	91,90%	93,80%
Multilayer Perceptron	93,70%	94,25%	93,70%

Tabela 9 - *Wisconsin Diagnostic Breast Cancer Database*

WDBC	base original (30 caract)	base PCA (15 comp)	base Redutos (16 caract)
JADE	93,94%	92,11%	92,99%
BayesNet	94,20%	90,80%	92,30%
Decision Table	94,20%	93,25%	95,65%
J48	92,95%	93,35%	94,55%
LibSVM	96,25%	88,95%	96,45%
Multilayer Perceptron	96,65%	95,40%	95,45%

Analisando nas tabelas os dados obtidos, vemos que o método baseado em seleção de características utilizado, *rough sets* [2], [3], [4], apresentou uma melhoria nos resultados de alguns classificadores (Tabela 9). Por outro lado, nem sempre reduz a dimensionalidade do espaço de características. O método de redução de características baseado em PCA [7], [8], também apresentou melhoria no resultado de alguns classificadores (Tabela 7, Tabela 8 e Tabela 9), mas os resultados são fortemente dependentes da distribuição dos dados.

## 6. CONCLUSÃO

Este documento apresentou o projeto intitulado “Desenvolvimento de Métodos para Redução de Dimensionalidade do Espaço de Características para Reconhecimento de Padrões”. Depois de um estudo geral sobre os métodos para redução de dimensionalidade do espaço de características [1], foram feitos estudos específicos sobre o uso do PCA [7], [8] e o uso do *Rough Sets* [2], [3], [4].

Após estes estudos, utilizando o método baseado em seleção de características (*Rough Sets*) e o método baseado em transformação de características (PCA), foram feitos testes em bases de dados públicas. Verificou-se que o método baseado em *Rough Sets* nem sempre apresenta redução de dimensionalidade, e que o método de transformação de características (PCA) depende da distribuição dos dados.

Os estudos realizados até o momento não são conclusivos com relação a indicar qual método é melhor indicado para cada base de dados.

## **7. AGRADECIMENTOS**

Agradecemos à FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) e a PROPP-UFU (Pró-Reitoria de Pesquisa e Pós-Graduação da Universidade Federal de Uberlândia) pelo apoio financeiro.

## **8. REFERÊNCIAS BIBLIOGRÁFICAS**

[1] L.J.P. VAN DER MAATEN. Dimensionality Reduction: A Comparative Review. Maastricht University, February 2008.

[2] PAWLAK, Z.; SKOWRON, A.. Rudiments of rough sets. Institute of Mathematics, Warsaw University, 177: 3–27, 2007.

[3] MARA, C.; CORREIA, C. S.. Rough Sets – Técnica de Redução de Atributos e Geração de Regras para Classificação de Dados. Campo Grande, 2005.

[4] SHEN, Q.; JENSEN, R. Rough Sets, Their Extensions and Applications. Department of Computer Science, The University of Wales, Aberystwyth, Ceredigion SY23 3DB, UK, July 2007.

[5] GULIATO, D.; SANTOS, J. C. S.. Granular Computing and Rough Sets to Generate Fuzzy Rules. Lecture Notes in Computer Science (LNCS), v. 5627, p. 317-326, July 2009.

[6] SANTOS, J. C. S.; GULIATO, D.. Proposta de um método para geração automática de regras fuzzy baseada na teoria dos rough sets. In: Latin American Informatics Conference, September 2009, Pelotas - RS.

[7] LINDSAY I SMITH. A tutorial on Principal Components Analysis. February 2002.

[8] CHRISTOPHER J.C. B. Geometric Methods for Feature Extraction and Dimensional Reduction. Redmond, November 2004.

[9] Rough Set Exploration System (RSES). Disponível: <http://alfa.mimuw.edu.pl/~rses/>: Acesso em: 08/06/2009.

[10] R-project. Disponível: <http://www.r-project.org/>: Acesso em: 07/10/2009.

[11] Weka. Disponível: <http://www.cs.waikato.ac.nz/ml/weka/>: Acesso em: 05/11/2009.