

# Seleção de Atributos Relevantes e Não Redundantes usando a Dimensão Fractal do Conjunto de Dados\*

Huei Diana Lee<sup>12</sup>, Maria Carolina Monard<sup>2</sup>, Feng Chung Wu<sup>13</sup>

<sup>1</sup>Laboratório de Bioinformática (LABI) – Universidade Estadual do Oeste do Paraná  
Caixa Postal 961, 85870-650 – Foz do Iguaçu, PR, Brasil

<sup>2</sup>Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo  
Caixa Postal 668, 13560-970 – São Carlos, SP, Brasil

<sup>3</sup>Instituto de Tecnologia em Automação e Informática (ITAI) – Foz do Iguaçu, PR, Brasil

huei@unioeste.br, mcmonard@icmc.usp.br, wufc@unioeste.br

**Abstract.** *Feature Selection is a central problem in machine learning, as non-relevant or redundant features may reduce accuracy and comprehensibility of hypothesis induced by supervised learning algorithms. Most of the state-of-art feature selection algorithms mainly focus on finding relevant features. However, it has been shown that relevance alone is not sufficient to select important features. Different approaches have been proposed to select features, among them the filter approach. This work proposes a filter that decouples relevance and redundancy analysis and introduces the use of Fractal Dimension to deal with redundant features. Empirical results on real world data show that Fractal Dimension is an appropriate criterion to filter out redundant features for supervised learning.*

**Resumo.** *Seleção de Atributos é um problema importante na área de aprendizado de máquina, pois atributos não relevantes ou redundantes podem reduzir a precisão e a compreensibilidade de hipóteses induzidas por algoritmos de aprendizado supervisionado. A maioria dos algoritmos selecionam, principalmente, atributos relevantes. Porém, tem sido observado que somente relevância não é suficiente para a seleção de atributos importantes. Dentre as diversas abordagens para seleção de atributos propostas, está a abordagem filtro. Neste trabalho é proposto um filtro que separa as análises de relevância e de redundância e introduz o conceito de Dimensão Fractal para tratar redundância de atributos. Resultados experimentais mostram que essa medida é um critério apropriado para filtrar atributos redundantes no aprendizado supervisionado.*

## 1. Introdução

A entrada para um algoritmo de aprendizado supervisionado consiste usualmente de um conjunto de  $N$  exemplos (ou casos) de treinamento  $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$  rotulados com os valores  $y$  de uma função  $f$  desconhecida  $y = f(\mathbf{x})$ , onde os valores  $\mathbf{x}_i$  são vetores

---

\*Trabalho desenvolvido com o apoio do Instituto de Tecnologia em Automação e Informática – ITAI e do Parque Tecnológico de Itaipu – PTL.

da forma  $\langle x_{i1}, x_{i2}, \dots, x_{iM} \rangle$  cujos componentes são valores discretos ou contínuos relacionados aos *atributos*  $A = \{A_1, A_2, \dots, A_M\}$ . Ou seja,  $x_{ij}$  denota o valor do atributo  $A_j$  do exemplo  $i$ . Dado esse conjunto de exemplos de treinamento, o algoritmo induz uma hipótese  $h$  que deve aproximar a verdadeira função  $f$ , tal que dados os valores  $\mathbf{x}$  de um novo exemplo,  $h$  prediz o valor  $y$  correspondente. No caso dos valores  $y$  pertencerem a um conjunto discreto de  $N_{C_i}$  classes, *i.e.*  $y \in \{C_1, \dots, C_{N_{C_i}}\}$ , a tarefa de aprendizado é chamada de *classificação*, tratada neste trabalho. No entanto, se  $y \in \mathfrak{R}$ , essa tarefa é denominada de *regressão*. O aprendizado de um classificador  $h$  é determinado pelo valor dos atributos. Ainda que teoricamente o uso de um maior número de atributos para descrever os exemplos deveria fornecer um maior poder de discriminação para aproximar  $f$ , isso pode não ocorrer, especialmente na presença de atributos irrelevantes e/ou redundantes, os quais, freqüentemente, confundem o algoritmo de aprendizado. Assim, a Seleção de Atributos – SA – é uma área de pesquisa há tempo explorada não somente em estatística mas também em Aprendizado de Máquina – AM – e Mineração de Dados – MD (Liu and Motoda, 1998). Os resultados obtidos, tanto teórica quanto experimentalmente, mostram que a SA melhora a predição de classificadores e reduz a complexidade do modelo  $h$ .

A meta da SA pode ser formalizada do seguinte modo (Yu and Liu, 2004): seja  $A' \subset A$  um subconjunto dos atributos de  $A$ , e  $f'(\mathbf{x}')$  os valores associados aos vetores correspondentes a  $A'$ . O objetivo da SA consiste em selecionar o subconjunto mínimo de atributos  $A'$  tal que  $\mathbf{P}(C|y = f'(\mathbf{x}')) \approx \mathbf{P}(C|y = f(\mathbf{x}))$ , onde  $\mathbf{P}(C|y = f'(\mathbf{x}'))$  e  $\mathbf{P}(C|y = f(\mathbf{x}))$  são as distribuições de probabilidades das  $N_{C_i}$  possíveis classes dados os valores dos atributos  $A'$  e  $A$  respectivamente. Esse subconjunto mínimo  $A'$  é denominado subconjunto *ótimo* de atributos. O seguinte exemplo, citado freqüentemente na literatura, ilustra esse conceito: considerando o conjunto  $A = \{A_1, A_2, A_3, A_4, A_5\}$  de atributos e  $y = f(A_1, A_2)$  uma função booleana, há somente oito possíveis exemplos tal que  $A_2 = \overline{A_3}$  e  $A_4 = \overline{A_5}$ . Assim, para determinar o conceito meta tem-se:  $A_1$  é indispensável;  $A_2$  ou  $A_3$ , mas não ambos, podem ser ignorados já que  $y = f(A_1, \overline{A_3})$ ;  $A_4$  e  $A_5$  podem ser ignorados. Nesse caso, existem dois subconjuntos  $A'$  ótimos,  $\{A_1, A_2\}$  e  $\{A_1, A_3\}$ , e a meta da SA é encontrar pelo menos um desses subconjuntos. Entretanto, o número de subconjuntos de atributos cresce exponencialmente com o número de atributos em  $A$  e encontrar o subconjunto ótimo de atributos pode ser NP (Kohavi and John, 1997).

Os diversos modelos de SA propostos na literatura podem ser categorizados nos modelos *wrapper* e filtro. O primeiro utiliza o próprio algoritmo de aprendizado para determinar, em cada iteração, a precisão do classificador  $h$  induzido utilizando o subconjunto de atributos selecionados nessa iteração. Ao final é considerado o melhor subconjunto de atributos aquele que melhora a precisão de  $h$  (Kohavi and John, 1997). A maior desvantagem dos métodos *wrapper* é que, além de serem específicos ao algoritmo considerado, são computacionalmente muito caros para conjuntos de exemplos descritos por um grande número de atributos. Diferentemente, o modelo filtro separa a SA do algoritmo de aprendizado que utilizará o subconjunto selecionado. A idéia é filtrar atributos irrelevantes, segundo algum critério, antes do aprendizado ocorrer.

Além dos atributos irrelevantes, tem sido observado que atributos *redundantes* também afetam a precisão dos classificadores induzidos e, portanto, deveriam ser eliminados (Koller and Sahami, 1996; Hall, 2000). Considera-se que dois atributos são redundantes entre si quando seus valores estão correlacionados, parcial ou completamente, tais

como os atributos  $A_2$  e  $A_3$  do exemplo previamente apresentado.

Em geral, os métodos de SA selecionam os atributos pela avaliação individual ou pela avaliação de subconjuntos de atributos. No caso de avaliação individual, os atributos são ordenados considerando a sua importância na discriminação das  $N_{C_i}$  classes. Esses métodos somente removem atributos irrelevantes pois espera-se que atributos redundantes tenham a mesma importância na discriminação das classes. Contudo, métodos que avaliam subconjuntos de atributos procurando por subconjuntos mínimos podem remover tanto atributos irrelevantes quanto redundantes. Assim, a maioria dos métodos existentes para a SA que tratam tanto relevância quanto redundância de atributos, o fazem de maneira implícita por meio da avaliação de subconjuntos de atributos. Ainda que esses métodos geralmente apresentem melhores resultados que os métodos que não lidam com a redundância de atributos, seu alto custo computacional pode torná-los ineficientes para conjuntos de dados com alta dimensionalidade. Recentemente foi proposto o uso da abordagem filtro considerando o modelo de tratamento da relevância e da redundância de atributos como dois procedimentos separados (Yu and Liu, 2004). A vantagem desse modelo sobre o modelo anterior é que, por meio da separação da análise de relevância e de redundância, permite encontrar um subconjunto que aproxima o subconjunto ótimo mas evita o custo computacional do modelo tradicional da busca por subconjuntos.

Neste trabalho investigamos o modelo proposto por Yu and Liu (2004) e propomos o uso da Dimensão Fractal –DF – como procedimento para tratar a redundância de atributos. Ainda que o conceito de DF seja frequentemente utilizado na detecção de agrupamento de dados e na indexação de estruturas de alta dimensionalidade, não é de nosso conhecimento que a DF tenha sido utilizada para realizar SA para algoritmos de aprendizado de máquina supervisionados, como proposto neste trabalho. Resultados experimentais obtidos com diversos conjuntos de dados utilizando diferentes procedimentos para tratar relevância de atributos e o algoritmo por nós proposto, que usa a DF para tratar redundância, são apresentados. Esses resultados mostram que a DF é um critério apropriado para tratar redundância de atributos.

O restante deste trabalho está organizado do seguinte modo: nas Seções 2 e 3 são apresentados brevemente conceitos sobre fractais e dimensão fractal. Na Seção 4 é descrito o algoritmo proposto neste trabalho. A configuração dos experimentos realizados, resultados e discussão são descritos na Seção 5. Conclusões e trabalhos futuros são apresentados na Seção 6.

## 2. Fractais

Fractais são definidos pela propriedade de auto-similaridade, ou seja, apresentam, parcial ou integralmente, as mesmas características para diferentes variações na escala em que estão sendo analisados. Assim, partes do fractal, o qual pode ser uma estrutura, um objeto ou um conjunto de dados, são similares, exata ou estatisticamente, ao fractal como um todo. Fractais possuem, em geral, características incomuns, por exemplo, o conhecido Triângulo de Sierpinsky – Figura 1. Ele não pode ser considerado um objeto Euclidiano unidimensional, pois possui perímetro

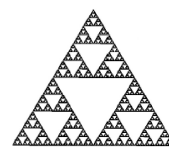


Figura 1: Triângulo de Sierpinsky

infinito, nem tão pouco um objeto Euclidiano bidimensional já que possui área nula. Dessa maneira, pode-se considerar uma dimensão fracionária, denominada de Dimensão Fractal (Mandelbrot, 1985).

Muitos dos conjuntos de dados reais comportam-se como fractais. Desse modo, torna-se natural a idéia de aplicar conceitos da teoria dos fractais para a análise desses conjuntos (Faloutsos and Kamel, 1994).

### 3. Dimensão Fractal de um Conjunto de Dados

A utilização do conceito de Dimensão Fractal – DF – está associada à idéia da existência de redundância nos conjuntos de dados e da possibilidade desses conjuntos serem bem aproximados em dimensões menores. A idéia principal é empregar a DF do conjunto de dados, a qual é relativamente não afetada por atributos redundantes, para determinar a quantidade e quais são os atributos não redundantes segundo o critério de DF (Sousa et al., 2002).

Pode-se definir, desse modo, as idéias de dimensão imersa e dimensão intrínseca. A primeira idéia corresponde à dimensão do espaço de endereçamento, ou seja, o número de atributos do conjunto de dados. Porém, o conjunto de dados pode estar representando um objeto que possui uma dimensão menor que a do espaço em que está imerso. Assim, a dimensão intrínseca é a dimensão espacial do objeto representado pelo conjunto de dados. Conceitualmente, se um conjunto de dados possui todas as suas variáveis (atributos) independentes umas das outras, então sua dimensão intrínseca será igual a sua dimensão imersa. Porém, toda vez que existir uma correlação entre duas ou mais variáveis, a dimensão intrínseca do conjunto de dados é reduzida de acordo. Usualmente, correlações entre os atributos ou a própria existência dessas correlações não é conhecida. Por meio da dimensão intrínseca do conjunto de dados é possível decidir quantos atributos são necessários para caracterizá-lo. Diferentes tipos de correlação podem reduzir a dimensão intrínseca em diferentes proporções, até mesmo em proporções fracionárias. Desse modo, pode-se utilizar o conceito de Dimensão Fractal como sendo a dimensão intrínseca do conjunto de dados (Traina et al., 2000).

Existem diversas medidas para a DF. Para fractais exatamente auto-similares, *i.e.* que podem ser caracterizados por meio de regras de construção bem definidas, a Dimensão Fractal é dada por  $D = \log(R)/\log(1/e)$ , onde  $R$  representa a quantidade de réplicas e  $1/e$  em que escala as réplicas são geradas a cada iteração. Para o exemplo do triângulo de Sierpinsky mencionado na Seção 2, a DF seria  $D = \log(3)/\log(2) = 1,58496$ , pois são geradas três réplicas em escala  $1:1/2$  a cada iteração (Figura 2).

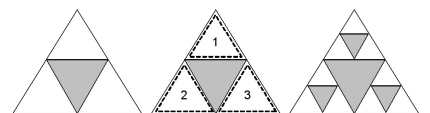


Figura 2: Construção do Triângulo de Sierpinsky

Para fractais estatisticamente auto-similares, como conjuntos de dados reais, uma das maneiras para a definição da DF é representada pela Dimensão Fractal de Correlação  $D_2$ , que pode ser calculada pelo método *Box Count Plot* (Faloutsos and Kamel, 1994). A idéia consiste, primeiramente, na construção de um reticulado sobre o conjunto de dados de células de lado  $r$ . Então conta-se o número de pontos dentro da  $i$ -ésima célula de tamanho  $r$ , denominado  $C_{r,i}$ . A Dimensão Fractal de Correlação  $D_2$  é definida como:  $D_2 = \frac{\partial \log(\sum_i C_{r,i}^2)}{\partial \log(r)}$ ,  $r \in [r_{min}, r_{max}]$ .

Em teoria, fractais exatamente auto-similares são infinitos. Na prática, conjuntos de dados reais, os quais possuem um número finito de pontos, são considerados fractais estatisticamente auto-similares para um determinado intervalo de escalas  $r \in (r_{min}, r_{max})$  se obedecem um regra de construção bem definida nesse intervalo. Desse modo, a dimensão intrínseca de um determinado conjunto de dados pode ser medida como o coeficiente angular da reta que melhor se ajusta ao trecho linear do gráfico em escala logarítmica de  $\sum_i C_{r,i}^2$  por  $r$  (Traina et al., 2000). Neste trabalho, o termo Dimensão Fractal de Correlação será simplesmente denominado de Dimensão Fractal.

#### 4. Descrição do Algoritmo Proposto

O algoritmo proposto neste trabalho para a seleção de atributos, denominado de *Fractal Dimension-Based Filter* – FDimBF, pertence à abordagem filtro e segue o modelo proposto por Yu and Liu (2004), ilustrado na Figura 3, o qual realiza a seleção de atributos em duas etapas: primeiramente é feita a análise de relevância para determinar o subconjunto de atributos relevantes em relação à classe, removendo os atributos irrelevantes; na segunda etapa, por meio da análise de redundância, são determinados e removidos os atributos redundantes a partir do subconjunto que contém apenas os atributos relevantes, produzindo o subconjunto final de atributos selecionados. O algoritmo de Yu and Liu (2004), *Fast Correlation-Based Filter* – FCBF, utiliza a medida *Symmetrical Uncertainty* (Press et al., 1992) como a medida de correlação para aproximar tanto a análise de relevância quanto a análise de redundância. O FCBF apresenta a vantagem, sobre as abordagens tradicionais para avaliação de subconjuntos de atributos, de que por meio da separação das tarefas de análise de relevância e redundância, ele ameniza o alto custo da busca por subconjuntos de atributos.

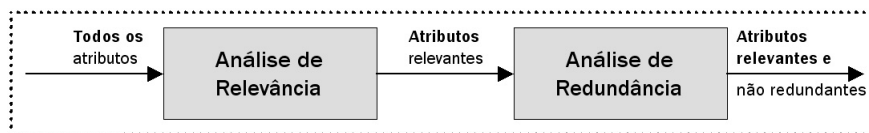


Figura 3: Modelo para Seleção de Atributos (Yu and Liu, 2004)

No algoritmo FDimBF, propomos o uso da Dimensão Fractal como medida para tratar a redundância de atributos. Para realizar a análise de relevância, propomos o uso de duas medidas: uma baseada em ganho de informação, algoritmo FDimBF(1), e outra baseada em distância, algoritmo FDimBF(2).

Especificamente, para realizar a análise de relevância em relação ao atributo classe usando a medida de ganho de informação, utilizamos o algoritmo  $\mathcal{C}4.5$  (Quinlan, 1993) e os atributos são classificados de acordo com o número de vezes que aparecem nas regras induzidas. Para medir a relevância dos atributos em relação à classe usando uma medida de distância, empregamos o algoritmo ReliefF (Robnik-Sikonja and Kononenko, 2003) para ordenar os atributos. Esse algoritmo procura pelos exemplos mais próximos da mesma classe e de classes diferentes e atribui pesos aos atributos de acordo com quão bem eles diferenciam esses exemplos. Esse processo é repetido  $m$  vezes. Em geral,  $m$  é definido em função do número de exemplos presentes no conjunto de dados.

Como mencionado anteriormente, para tratar a análise de redundância, neste trabalho propomos a utilização da DF. Para medir a DF dos conjuntos de dados, foi empregada a ferramenta *Measure Distance Exponent* – MDE (Traina et al., 2003). Atributos

redundantes, considerando a Dimensão Fractal, podem ser definidos como aqueles que quando excluídos do conjunto de dados não causam uma modificação no valor da DF recalculada. O método usado pelo MDE consiste na medição do valor da DF,  $D$ , a partir do conjunto de dados original e do valor da Dimensão Fractal Parcial,  $pD$ , ignorando um atributo por vez. Em outras palavras a  $pD$  é calculada tomando-se em consideração todos os atributos exceto o  $i$ -ésimo atributo sob observação. O processo continua selecionando o atributo que permite a diferença mínima entre  $D$  e  $pD$ . Se a diferença é menor que um limiar mínimo, o qual determina quão preciso o conjunto de dados, descrito por apenas os atributos selecionados, precisa ser para preservar as características do conjunto de dados original, esse atributo pode ser considerado como de contribuição pequena para a caracterização do conjunto de dados original. Esse processo continua, considerando o restante dos atributos e fazendo com que  $D = pD$  e aplicando o procedimento descrito, até que não existam mais atributos a serem removidos. Ao final do processo, os atributos estarão inversamente ordenados de acordo com sua contribuição, em termos de redundância, para a medição da DF do conjunto de dados (Traina et al., 2000).

## 5. Experimentos Realizados

Nesta seção são apresentados os conjuntos de dados e a configuração dos experimentos realizados. São também apresentados os resultados obtidos e a discussão desses resultados.

### 5.1. Descrição dos Conjuntos de Dados

Foram utilizados três conjuntos de dados obtidos do Repositório de Dados UCI (Blake et al., 1998): Bupa, Pima e Breast Cancer, brevemente descritos a seguir.

**Breast Cancer:** o problema é prever se uma amostra de tecido de mama obtida de uma paciente é maligna ou benigna baseada em dados histológicos.

**Pima:** o problema é prever se uma paciente, mulher de descendência indígena Pima com idade mínima de 21 anos, seria classificada como diabética, segundo o critério estabelecido pela Organização Mundial de Saúde, fornecendo dados clínicos e laboratoriais.

**Bupa:** o problema é prever se um paciente, do sexo masculino, possui ou não disfunção hepática tomando-se como base diversos exames sanguíneos e a quantidade de álcool consumida.

A Tabela 1 mostra um resumo das características desses três conjuntos de dados. Para cada um deles são apresentados: número de exemplos, número e percentagem de exemplos duplicados (aparecem mais de uma vez) ou conflitantes (mesmos valores para todos os atributos, porém pertencentes a classes diferentes), número total de atributos, número de atributos contínuos e nominais (discretos), valores e distribuição das classes, erro cometido no caso de novos exemplos serem classificados como sendo pertencentes à classe majoritária – CM – e existência ou não de valores desconhecidos (? – última coluna).

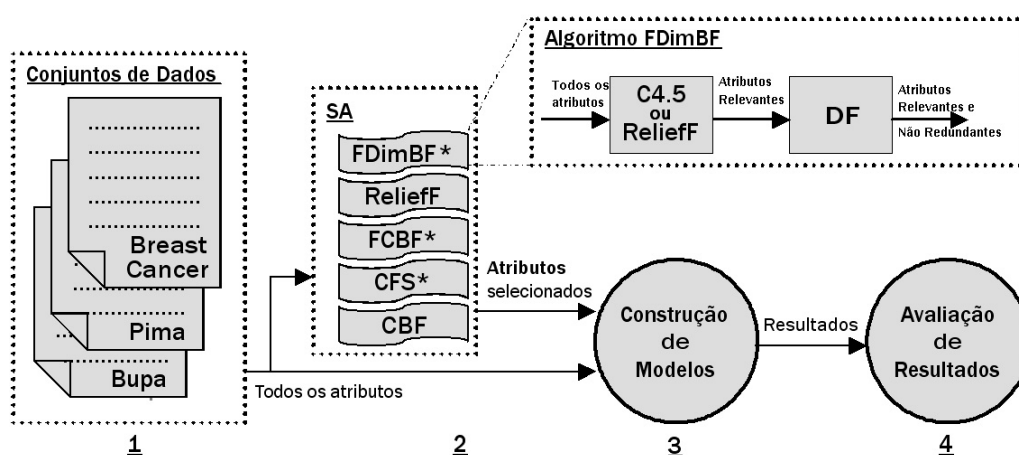
### 5.2. Configuração dos Experimentos

Os experimentos foram organizados em quatro etapas ilustradas na Figura 4.

Na primeira etapa, foram realizadas a limpeza e a preparação dos dados. Dos três conjuntos de dados considerados neste trabalho, apenas o conjunto de dados Breast Cancer teve que ser submetido à limpeza de dados, resultando em 683 exemplos descritos pelo

Conjunto de Dados	# Exemplos	# Duplicados ou Conflitantes (%)	# Atributos (cont.,nom.)	Classes	Classe %	Erro da CM	?
Breast Cancer	699	8 (1.15%)	9 (9,0)	2	65.52%	34.48% sobre 2	Sim
				4	34.48%		
Pima	769	1 (0.13%)	8 (8,0)	0	65.02%	34.98% sobre 0	Não
				1	34.98%		
Bupa	345	4 (1.16%)	6 (6,0)	1	42.03%	42.03% sobre 2	Não
				2	57.97%		

**Tabela 1: Resumo dos Conjuntos de Dados**



**Figura 4: Configuração dos Experimentos**

mesmo número de atributos. Essa tarefa consistiu na remoção de valores desconhecidos da seguinte maneira: quando os valores desconhecidos encontravam-se concentrados em alguns poucos exemplos, esses exemplos eram removidos, porém quando os valores desconhecidos estavam concentrados em um atributo, a coluna correspondente era removida do conjunto de dados. A principal razão para a remoção de valores desconhecidos do conjunto de dados é que alguns dos algoritmos utilizados nesses experimentos tratam valores faltantes de modo especial (Batista and Monard, 2003), enquanto outros algoritmos não tratam esse tipo de informação. Assim, com o intuito de não introduzir interferências associadas ao uso de um ou outro método para tratar esse problema, foi decidida a remoção de valores desconhecidos do conjunto de dados. Ao final dessa etapa, os dados foram transformados para a sintaxe requerida por cada um dos algoritmos e ferramentas utilizados neste trabalho.

Na segunda etapa, subconjuntos de atributos foram selecionados utilizando algoritmos frequentemente utilizados na abordagem filtro. Além do algoritmo FDimBF proposto, foram considerados quatro algoritmos para a seleção de atributos. Dois desses algoritmos, ReliefF (Robnik-Sikonja and Kononenko, 2003) e FCBF (Yu and Liu, 2004), citados anteriormente, avaliam atributos individualmente. Os algoritmos CFS (*Correlation-based Feature Selection*) (Hall, 2000) e CBF (*Consistency-based Filter*) (Liu and Setiono, 1996) avaliam subconjuntos de atributos. Como mencionado, o algoritmo ReliefF busca por exemplos vizinhos de diferentes classes e atribui pesos aos atributos de acordo com quão bem eles diferenciam esses exemplos. Já o algoritmo CFS avalia a importância de um subconjunto de atributos baseado na habilidade preditiva individual de cada atributo

e o grau de correlação entre esses atributos. CBF avalia os subconjuntos de atributos de acordo com sua consistência em relação à classe buscando por combinações de atributos cujos valores particionem os dados em subconjuntos com alguma classe majoritária. O algoritmo FCBF realiza a seleção de atributos em duas etapas: primeiramente realiza a análise de relevância para determinar o subconjunto de atributos relevantes em relação à classe, removendo os atributos irrelevantes; na segunda etapa, por meio da análise de redundância, determina e remove os atributos redundantes a partir do subconjunto que contém apenas os atributos relevantes, produzindo o subconjunto final de atributos selecionados. Todos esses algoritmos, a exceção do algoritmo proposto neste trabalho, estão implementados na ferramenta Weka (Witten and Frank, 1999) e foram executados considerando seus parâmetros configurados com os valores padrões. Note que os algoritmos marcados com \* na Figura 4 são aqueles que tratam tanto o problema da relevância de atributos, em relação ao atributo classe, quanto o problema da redundância de atributos.

Na terceira etapa, modelos (classificadores) foram induzidos usando todos os atributos remanescentes da etapa 1 e apenas os atributos selecionados na etapa anterior. Esses modelos foram construídos utilizando o algoritmo C4.5.

Durante a última etapa, os resultados foram avaliados por meio da estimativa da média do erro de cada um dos modelos construídos empregando validação cruzada com 10 partições (*10 fold cross-validation*). Esse modo de avaliação foi escolhido pois, para conjuntos de dados reais, conhecimento prévio sobre que atributos são importantes, em geral, não está disponível. Desse modo, a precisão preditiva é comumente utilizada como uma medida indireta para avaliar a qualidade dos atributos selecionados.

### 5.3. Resultados e Discussão

Para cada conjunto de dados, foram selecionados subconjuntos de atributos usando as duas versões do algoritmo proposto neste trabalho, *i.e.* FDimBF(1) e FDimBF(2), e os algoritmos ReliefF, CFS, CBF e FCBF, totalizando 18 experimentos. Os resultados dos experimentos, considerando os erros dos classificadores estimados por meio de validação cruzada com 10 partições, foram comparados usando ANOVA com fator único. Para todos os três conjuntos de dados não houve diferença estatisticamente significativa para um nível de significância de 95% (Bupa:  $p = 0,7064$ ; Pima:  $p = 0,9020$  e Breast Cancer:  $p = 0,9239$ ). Isso indica que, para esses conjuntos de dados, os algoritmos para SA utilizados neste trabalho apresentam, estatisticamente, desempenho semelhante quanto à precisão do modelo construído.

Por questões de espaço, apenas a tabela completa contendo os resultados obtidos nos experimentos para o conjunto de dados Breast Cancer é apresentada – Tabela 2. Nessa tabela, na primeira coluna são mostrados os atributos de cada conjunto de dados; nas colunas 2 a 8, para cada uma das abordagens utilizadas neste trabalho, é mostrado se o atributo foi ou não selecionado, a porcentagem de atributos selecionados por aquela abordagem (% Sel.), o erro dos classificadores estimados por meio de validação cruzada com 10 partições (Erro VC) e o erro padrão (SE). Nas últimas colunas, #Ocor. e %Ocor., são apresentados o número de vezes e a respectiva porcentagem que cada atributo foi selecionado, considerando todos os algoritmos utilizados.

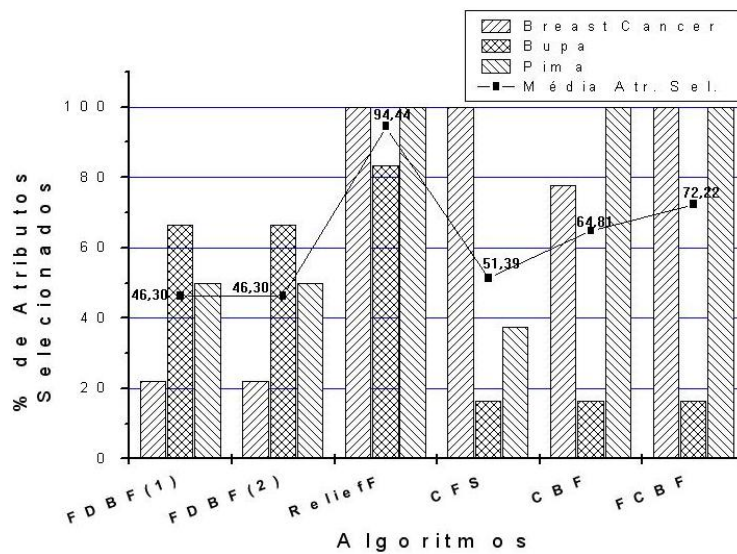
Na Figura 5 são apresentados os resultados obtidos com todos os conjuntos de dados. O gráfico de barras mostra a porcentagem de atributos selecionados por cada algoritmo no respectivo conjunto de dados. O gráfico de linhas e pontos mostra a média



Breast Cancer	Sem SA	FDimBF(1)	FDimBF(2)	ReliefF	CFS	CBF	FCBF	# Ocor.	% Ocor.
1	◇	◇	◇	◇	◇	◇	◇	6	100,00
2	◇			◇	◇		◇	4	66,67
3	◇	◇		◇	◇		◇	4	66,67
4	◇			◇	◇	◇	◇	4	66,67
5	◇			◇	◇	◇	◇	4	66,67
6	◇			◇	◇	◇	◇	4	66,67
7	◇		◇	◇	◇	◇	◇	5	83,33
8	◇			◇	◇	◇	◇	4	66,67
9	◇			◇	◇		◇	3	50,00
% Sel.	100,00	22,22	22,22	100,00	100,00	77,78	100,00		
Erro VC	4,10	5,70	5,30	4,10	4,10	4,40	4,10		
SE	1,00	0,70	0,70	1,00	1,00	0,50	1,00		

**Tabela 2: Resultados Obtidos dos Experimentos: Breast Cancer**

total, *i.e.* considerando todos os conjunto de dados, de atributos selecionados por cada algoritmo. Pode ser observado que as duas versões do algoritmo FDimBF proposto, juntamente com o algoritmo CFS selecionaram, em média, uma quantidade menor de atributos quando comparada aos outros algoritmos (46,30% e 51,39% respectivamente). Também, é interessante ressaltar que os subconjuntos de atributos selecionados pelos algoritmos FDimBF e CFS contém os atributos mais freqüentemente selecionados pelos algoritmos considerados.



**Figura 5: Percentagem de Atributos Selecionados por cada Algoritmo**

## 6. Conclusões

Neste trabalho investigamos um novo modelo de Seleção de Atributos, no qual o tratamento de relevância e redundância de atributos é realizado separadamente, e propomos um algoritmo, FDimBF, que usa a Dimensão Fractal para tratar a redundância. Duas versões desse algoritmo, as quais consideram dois critérios diferentes para selecionar atributos relevantes, foram comparadas experimentalmente com outros algoritmos de SA propostos na literatura. Os resultados obtidos são encorajadores pois mostram que a DF é um critério apropriado para tratar redundância de atributos.

Como trabalhos futuros pretendemos aplicar o algoritmo proposto em bases de dados reais de maior dimensionalidade, da área médica, nas quais estamos trabalhando

na extração de conhecimento com o auxílio de especialistas do domínio, e investigar a qualidade dos atributos selecionados por esse algoritmo também do ponto de vista do especialista.

## Referências

- Batista, G. E. A. P. A. and Monard, M. C. (2003). An Analysis of Four Missing Data Treatment Methods for Supervised Learning. *Applied Artificial Intelligence*, 17(5):519–533.
- Blake, C., Keogh, E., and Merz, C. (1998). UCI repository of machine learning databases. <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- Faloutsos, C. and Kamel, I. (1994). Beyond uniformity and independence: Analysis of r-trees using the concept of fractal dimension. In *Proceedings of the 13th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pages 4–13, Minneapolis, MN.
- Hall, M. A. (2000). Correlation-based feature selection for discrete and numeric class machine learning. In *Proc. 17th International Conf. on Machine Learning*, pages 359–366, San Francisco, CA. Morgan Kaufmann.
- Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, pages 273–324.
- Koller, D. and Sahami, M. (1996). Toward optimal feature selection. In *Proceedings of the Thirteenth International Conference on Machine Learning*, pages 284–292, Bari, Italy.
- Liu, H. and Motoda, H. (1998). *Feature Selection for Knowledge and Data Mining*. Kluwer Academic Publishers, Massachusetts.
- Liu, H. and Setiono, R. (1996). A probabilistic approach to feature selection – a filter solution. In *Proceedings of the Thirteenth International Conference on Machine Learning*, pages 319–327, Bari, Italy.
- Mandelbrot, B. B. (1985). *The Fractal Geometry of Nature: Updated and Augmented*. W. H. Freeman and Company, New York.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press.
- Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann. San Francisco, CA.
- Robnik-Sikonja, M. and Kononenko, I. (2003). Theoretical and empirical analysis of ReliefF and RReliefF. *Mach. Learn.*, 53(1-2):23–69.
- Sousa, E. P. M., Traina, C., Traina, A. J. M., and Faloutsos, C. (2002). How to use fractal dimension to find correlations between attributes. In *Workshop Notes of KDD 2002 Workshop on Fractals and Self-similarity in Data Mining: Issues and Approaches*, pages 26–30, Edmonton, Canada.
- Traina, C., Traina, A. J. M., and Faloutsos, C. (2003). Mde - measure distance exponent manual. (Internal Document).
- Traina, C., Traina, A. J. M., Wu, L., and Faloutsos, C. (2000). Fast feature selection using fractal dimension. *XV Brazilian Data Base Symposium*, pages 158–171.
- Witten, I. H. and Frank, E. (1999). *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, volume 1. Morgan Kaufmann.
- Yu, L. and Liu, H. (2004). Efficient feature selection via analysis of relevance and redundancy. *Journal of Machine Learning Research*, 5:1205–1224.