

Seleção de Atributos para Algoritmos de Aprendizado de Máquina Supervisionado Utilizando como Filtro a Dimensão Fractal

Huei Diana Lee^{1 2}

Maria Carolina Monard²

¹ Universidade Estadual do Oeste do Paraná
LABI - Laboratório de Bioinformática
Caixa Postal 961, 85870-650 - Foz do Iguaçu, PR, Brasil
e-mail: huei@unioeste.br

² Universidade de São Paulo
Instituto de Ciências Matemáticas e de Computação
Caixa Postal 668, 13560-970 - São Carlos, SP, Brasil
e-mail: mcmonard@icmc.sc.usp.br

Resumo: A Seleção de um Subconjunto de Atributos é um problema importante para a área de Aprendizado de Máquina, na qual o algoritmo de aprendizado enfrenta o problema de selecionar um subconjunto de atributos sobre o qual focar sua atenção, enquanto ignora o restante deles. Vários métodos, os quais podem ser agrupados em *embedded*, filtro e *wrapper*, têm sido propostos na literatura para abordar esse problema. Neste trabalho é proposto um filtro que utiliza a Dimensão Fractal como critério de relevância para selecionar um subconjunto de atributos. Resultados experimentais obtidos com diversos conjuntos de dados utilizando o critério de Dimensão Fractal bem como outros critérios freqüentemente utilizados na abordagem filtro são também apresentados. O panorama geral dos resultados obtidos mostra que, com performance similar, o critério de filtragem utilizando a Dimensão Fractal permitiu, na maioria das vezes, a seleção de um subconjunto de atributos menor que o selecionado utilizando outros critérios, tais como *Column Importance* e *C4.5*.

Palavras-Chaves: Seleção de Atributos; Dimensão Fractal; Aprendizado de Máquina.

1 Introdução

Em Aprendizado de Máquina – AM – supervisionado, a um algoritmo de indução é apresentado, tipicamente, um conjunto de exemplos de treinamento, no qual cada caso é descrito por um vetor de valores de atributos e um rótulo para a classe. A tarefa do algoritmo de aprendizado é induzir um classificador cujo objetivo é rotular, com uma boa precisão, novos casos [Mitchell, 1997]. Um dos problemas centrais em AM é a Seleção de um Subconjunto de Atributos – SSA (*Feature Subset Selection*), em que o algoritmo de aprendizado enfrenta o problema de selecionar um subconjunto de atributos sobre o qual focar sua atenção, enquanto ignora o restante deles [Kohavi and John, 1997].

Existem várias maneiras de realizar SSA [Liu and Motoda, 1998]. Neste trabalho é proposto o uso da abordagem filtro considerando-se como critério de relevância a Dimensão Fractal do conjunto de dados. Algumas das aplicações que utilizam a teoria de fractais incluem a determinação de estrutura de indexação de alta dimensionalidade e a detecção de agrupamentos. Porém, a teoria dos Fractais é ainda pouco utilizada para o problema da Seleção de um Subconjunto de Atributos para algoritmos de aprendizado. Diversos experimentos utilizando algumas bases de dados médicas e o critério de Dimensão Fractal para realizar SSA são apresentados. Os resultados desses experimentos são comparados a resultados obtidos por outros critérios comumente utilizados como filtros.

Nas próximas seções são descritos os problemas da seleção de atributos e, mais especificamente, SSA empregando-se o conceito de Dimensão Fractal (Seções 2, 3 e 4). Nas Seções 5 e 6 são descritos os conjuntos de dados e as ferramentas utilizados nos experimentos; na Seção 7 é especificada a configuração desses experimentos e na Seção 8 os resultados obtidos. Algumas considerações finais e trabalhos futuros encontram-se na Seção 9.

2 Seleção de Atributos

O processo de seleção de atributos de um conjunto de exemplos, a ser utilizado por algoritmos de AM, tem como objetivo escolher o melhor subconjunto de atributos, segundo algum critério em particular, que representa bem esses exemplos. Esse processo pode ser realizado, basicamente, de duas maneiras [Guyon and Elisseeff, 2003, Lee, 2000]: Seleção de um Subconjunto de Atributos, tratado neste trabalho, no qual o objetivo é diminuir o espaço de descrição do conjunto de exemplos, e Construção de Atributos.

Dentre os modos como a SSA pode ser realizada, pode-se citar as abordagens *embedded*, filtro e *wrapper* [John et al., 1994]. Na primeira abordagem, a SSA é realizada internamente pelo próprio algoritmo, *i.e.*, está embutida no algoritmo de indução.

A abordagem filtro introduz um processo separado, o qual ocorre antes da aplicação do algoritmo de indução propriamente dito. A idéia é filtrar atributos irrelevantes, segundo algum critério, antes da indução ocorrer. Esse passo de pré-processamento considera características gerais do conjunto de dados para selecionar alguns atributos e excluir outros. Sendo assim, métodos de filtros são independentes do algoritmo de indução que, simplesmente, receberá como entrada a saída, ou seja, o conjunto dos exemplos descritos em uma dimensão menor, fornecida pelo filtro.

A terceira abordagem, *wrapper*, também ocorre externamente ao algoritmo básico de indução, porém utilizando tal algoritmo como uma caixa preta para analisar o conjunto de atributos fornecido. Em outras palavras, métodos *wrapper* geram um subconjunto candidato de atributos, executam o algoritmo de indução considerando apenas esse subconjunto selecionado do conjunto de treinamento, e utilizam a precisão resultante do classificador construído para avaliar o subconjunto de atributos em questão. Esse processo é repetido para cada subconjunto de atributos até que o critério de parada determinado seja satisfeito. Deve ser observado que, contrariamente à abordagem filtro, a abordagem *wrapper* é computacionalmente cara.

Neste trabalho é considerada a abordagem filtro utilizando-se como critério de importância a Dimensão Fractal do conjunto de dados.

3 Fractais

Os fractais são definidos pela propriedade de auto-similaridade, ou seja, apresentam as mesmas características para diferentes variações em escala e tamanho. 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.

Os fractais possuem, em geral, características incomuns, por exemplo, o conhecido Triângulo de Sierpinsky – Figura 1, possui perímetro infinito e área nula. Portanto, não pode ser considerado um objeto Euclidiano unidimensional, pois possui perímetro infinito, nem tão pouco um objeto Euclidiano bi-dimensional já que possui área nula [Sousa, 2003]. Dessa maneira, pode-se considerar uma dimensão fracionária, denominada de Dimensão Fractal [Mandelbrot, 1985].

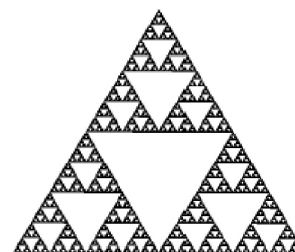


Figura 1: Exemplo de Fractal - Triângulo de Sierpinsky

Muitos dos conjuntos de dados reais comportam-se como fractais [Schroeder, 1991]. Desse modo, torna-se natural a idéia de aplicar conceitos da teoria dos fractais para a análise desses conjuntos [Faloutsos and Kamel, 1994].

4 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, como critério para determinar quantos e quais são os atributos mais

relevantes.

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 são conhecidas. 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 dimensão fractal. 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 $D = \log(R)/\log(e)$, onde R representa a quantidade de réplicas e e em que escala as réplicas são geradas a cada interação. Para o exemplo do triângulo de Sierpinsky mencionado na Seção 3, a Dimensão Fractal seria $D = \log(3)/\log(2) = 1,58496$, pois são geradas três réplicas em escala 1:2 a cada interação, como mostrado na Figura 2.

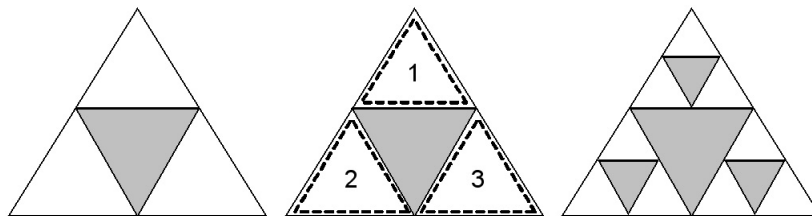


Figura 2: Exemplo de Iteração para Construção do Triângulo de Sierpinsky

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 lado 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 uma 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.

5 Descrição dos Conjuntos de Dados

Foram utilizados para a realização dos experimentos quatro conjuntos de dados obtidos do Repositório de Dados da UCI [Blake et al., 1998]: Bupa, Pima, Breast Cancer e Hepatitis, os quais são brevemente descritos a seguir.

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.

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

Breast Cancer O problema é prever se uma amostra de tecido de mama obtida de uma paciente é maligna ou benigna.

Hungarian O problema é prever se um paciente tem ou não doenças cardíacas.

A Tabela 1 mostra um resumo das características desses quatro 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.

Conjunto de Dados	# Exemplos	# Duplicados ou Conflitantes (%)	# Atributos (cont.,nom.)	Classes	Classe %	Erro da CM	Valores Desconhecidos
Bupa	345	4 (1.16%)	6 (6,0)	1	42.03%	42.03% sobre o valor 2	Não
				2	57.97%		
Pima	769	1 (0.13%)	8 (8,0)	0	65.02%	34.98% sobre o valor 0	Não
				1	34.98%		
Breast Cancer	699	8 (1.15%)	9 (9,0)	2	65.52%	34.48% sobre o valor 2	Sim
				4	34.48%		
Hungarian	294	1 (0.34%)	13 (13,0)	1	36.05%	36.05% sobre o valor 0	Sim
				0	63.95%		

Tabela 1: Resumo dos Conjuntos de Dados

6 Ferramentas Utilizadas

Dois ferramentas foram utilizadas para a realização dos experimentos, o *See5* e o *Measure Distance Exponent* – MDE. O *See5* (<http://www.rulequest.com>) é a versão comercial do algoritmo de aprendizado supervisionado *C4.5* [Quinlan, 1993], que induz árvores e regras de decisão. O MDE [Traina et al., 2003] permite que uma série de medidas sejam realizadas sobre um conjunto de dados, dentre elas o cálculo da DF de um conjunto de dados e a classificação dos atributos, segundo sua importância para a medida desse valor. Como entrada para o MDE é preciso fornecer um arquivo, no qual são descritos os dados seguidos do conjunto de exemplos. O MDE fornece como saída, além dos valores da DF e a classificação dos atributos, um gráfico mostrando o comportamento do conjunto de dados utilizando como base para o cálculo da DF o método *Box Count Plot*. A fim de ilustrar, a Figura 3 mostra a saída gráfica da ferramenta MDE para o conjunto de dados Bupa descrito originariamente por 6 (seis) atributos — Tabela 1.

A curva (a) representa o comportamento do conjunto de dados quando o gráfico logarítmico da somatória do número de pontos dentro de cada célula versus o tamanho da célula é plotado. As retas, (b) e (c), representam duas aproximações para o cálculo da DF. A primeira se ajusta à curva com o menor erro possível, enquanto a segunda reta é ajustada considerando um erro mínimo de 0,015 e um mínimo de cinco pontos na curva. Desse modo, deve-se escolher a reta que melhor se ajustou ao trecho linear da curva. No exemplo do gráfico apresentado na Figura 3, o melhor ajuste é dado pela reta (b) e portanto o valor de Dimensão Fractal que deve ser considerado é de 3,0871. Ou seja, é necessário o número máximo de de 4 (quatro) atributos para caracterizar esse conjunto de dados segundo a DF. O próximo passo consiste em determinar quais são esses quatro atributos.

De uma maneira simplificada, a classificação de que atributos são importantes, segundo a DF, é realizada por meio da determinação de que atributos, quando retirados do conjunto de dados, causam uma mudança significativa no valor da DF recalculada.

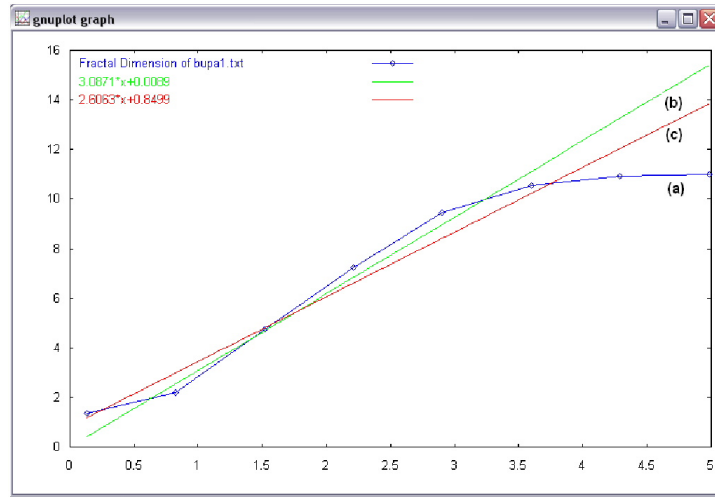


Figura 3: Gráfico Gerado pelo MDE Utilizando o Método *Box Count Plot*

A idéia consiste em medir a DF D do conjunto de dados como um todo e também as DFs parciais pD * desconsiderando um dos atributos de cada vez. O processo continua selecionando-se o atributo que permite a diferença mínima entre a pD e a D . Se essa diferença está dentro de um pequeno limiar, o qual determina o quão preciso o conjunto de dados resultante necessita ser para preservar as características do conjunto de dados original, pode-se considerar que esse atributo contribui minimamente para a caracterização do conjunto de dados original. O processo continua, considerando o conjunto de atributos restantes e a D como sendo igual a pD e repetindo o procedimento descrito, até que não existam mais atributos a serem removidos. Ao final do processo, os atributos estarão inversamente ordenados segundo sua importância para a medição da DF do conjunto de dados [Traina et al., 2000].

7 Experimentos Realizados

Os experimentos foram realizados em quatro fases como mostra a Figura 4.

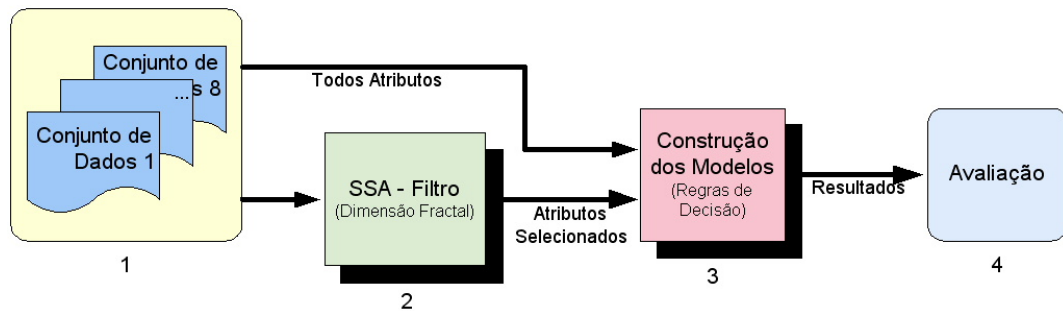


Figura 4: Configuração dos Experimentos

1. Pré-processamento dos Dados: essa fase consistiu de duas tarefas, a limpeza e a preparação dos dados. Na primeira delas, limpeza de dados, foram removidos os valores desconhecidos, o que implicou na remoção do atributo (coluna) que apresentava algum valor desconhecido. Entretanto, em alguns casos, os valores desconhecidos encontram-se concentrados em poucos exemplos. Nesses casos, o procedimento padrão consiste em remover primeiro esses exemplos (linhas) antes de reduzir a dimensão do conjunto de dados. Essa tarefa foi realizada somente para os conjuntos de dados Breast Cancer (resultando em 683 casos e mesmo número de atributos) e Hungarian (resultando em 246 casos e 10 atributos), pois

*DF parcial pD_i é calculada utilizando-se todos os atributos exceto o atributo i em questão.

apenas esses dois conjuntos possuíam valores desconhecidos. Essa abordagem foi escolhida para suprimir qualquer interferência relacionada aos diferentes mecanismos implementados nas ferramentas para tratar esse tipo de valores. Durante a segunda tarefa, preparação dos dados, duas etapas foram realizadas. Primeiramente, cada um dos conjuntos de dados deu origem a outros dois: um contendo o atributo Classe e outro desconsiderando esse atributo. É interessante notar que para o cálculo da Dimensão Fractal, todos os atributos, inclusive o atributo meta, são considerados indistintamente. Por essa razão foram criados esses oito conjuntos de dados distintos, com o objetivo de verificar a influência do atributo classe no cálculo da DF do conjunto de dados. Posteriormente, foram construídos os arquivos de entrada na sintaxe requerida pelas ferramentas utilizadas nos experimentos;

2. Seleção de um Subconjunto de Atributos: nessa fase foi utilizada a abordagem filtro considerando a Dimensão Fractal do conjunto de dados como referência para determinar quantos e quais atributos são importantes para a descrição do conjunto de dados. Nesse passo foi utilizada a ferramenta MDE;

3. Construção dos Modelos: considerando-se todos os atributos e o conjunto de atributos selecionados na fase anterior, foram induzidos os modelos, para cada um dos oito conjuntos de dados, utilizando-se a ferramenta *See5*;

4. Avaliação dos Resultados: para cada um dos modelos induzidos foi medido o erro utilizando-se 10 *fold cross-validation*. Os resultados obtidos nesses experimentos que usam a DF foram comparados com resultados obtidos previamente utilizando outros filtros.

8 Avaliação dos Resultados

Na Tabela 2 são apresentados os resultados obtidos neste trabalho e parte dos resultados obtidos em um trabalho anterior [Lee, 2000], no qual são considerados como critério de importância para a abordagem filtro a medida *Column Importance – CI* – disponibilizada pela ferramenta MineSetTM da Silicon Graphics Inc. e o algoritmo *C4.5*. A primeira coluna da Tabela 2 identifica o conjunto de dados e, entre parênteses, o erro da classe majoritária. Nas colunas 2, 3, 4 e 5 são apresentados as médias dos erros e os erros padrão dos modelos induzidos por *See5*, medidos utilizando-se 10 *fold cross-validation*, usando todos os atributos e somente os atributos filtrados considerando-se a DF, CI e *C4.5*, respectivamente. Nessas colunas, a percentagem de atributos selecionados é também mostrada entre parênteses. É apresentada apenas uma coluna para a DF, pois, como esperado, para os experimentos realizados foram selecionados exatamente o mesmo subconjunto de atributos considerado-se DF1 e DF2 (DF1: com o atributo Classe e DF2: sem o atributo Classe). Como pode ser observado, independentemente do número de atributos selecionados usando os três filtros, o erro dos classificadores induzidos utilizando a DF para selecionar os atributos, é competitivo com os outros dois filtros, exceto para o conjunto de dados Hungarian.

Conjunto de Dados (Erro da CM)	Todos Atributos	DF	CI	<i>C4.5</i>
Bupa (42,03)	31,90±2,00 (100,00%)	33,10±2,80 (66,67%)	41,42±2,85 (16,67%)	32,70±2,79 (100,00%)
Pima (34,98)	25,40±1,10 (100,00%)	27,80±2,00 (50,00%)	26,53±0,73 (75,00%)	25,88±0,99 (87,50%)
Breast Cancer (34,48)	5,00±1,20 (100,00%)	5,30±1,10 (20,00%)	5,86±0,84 (90,00%)	6,01±0,76 (80,00%)
Hungarian (36,05)	20,00±2,10 (100,00%)	31,80±2,80 (23,08%)	19,74±2,50 (76,92%)	20,09±2,59 (84,62%)

Tabela 2: Médias de Erros, Erros Padrão e Percentagem de Atributos Selecionados.

Na Tabela 3 são mostrados os resultados do Teste-*t* relacionado aos erros dos classificadores induzidos por *See5* após a seleção de atributos com os três filtros. Nas colunas 2 e 3 dessa tabela são identificados, respectivamente, o filtro e a percentagem de atributos selecionados. A coluna *t* mostra o valor do Teste-*t* (diferença significativa com grau de 95% para $|t| \geq 2,1$) considerando-se sempre a abordagem que utiliza a Dimensão Fractal comparada ao CI e ao *C4.5*. Como anteriormente, a SSA é denominada apenas DF pois, como mencionado, o mesmo subconjunto de atributos foi selecionado com e sem o atributo classe.

Conjunto de Dados	SSA	% Atributos Selecionados	t
Bupa	DF	66,67	
	CI	16,67	-0,93
	$\mathcal{C}4.5$	100,00	0,05
Pima	DF	50,00	
	CI	75,00	0,27
	$\mathcal{C}4.5$	87,50	0,38
Breast Cancer	DF	20,00	
	CI	90,00	-0,18
	$\mathcal{C}4.5$	80,00	-0,24
Hungarian	DF	23,08	
	CI	76,92	1,44
	$\mathcal{C}4.5$	84,62	1,37

Tabela 3: Comparação Entre os Resultados

de atributos, um número bem menor em relação a quantidade de atributos selecionados por CI (76,92%) e $\mathcal{C}4.5$ (84,62%).

O conjunto de dados para o qual a DF mostrou-se muito apropriada é o Breast Cancer. Para esse conjunto, além do erro do classificador ser competitivo com os erros obtidos utilizando os outros dois filtros, o número de atributos relevantes segundo a DF é muito menor que a quantidade selecionada com os outros filtros. Apesar do resultado obtido com o conjunto Hungarian, deve ser observado que em nenhum caso houve diferença significativa, com grau de 95%, entre os erros dos classificadores induzidos utilizando os três critérios de filtragem, o que mostra o potencial do uso da Dimensão Fractal como critério para filtrar atributos para serem utilizados por algoritmos de AM.

É interessante notar que os subconjuntos de atributos selecionados usando-se DF1 (com o atributo classe) e DF2 (sem o atributo classe) foram exatamente os mesmos. Esse resultado é esperado, pois significa que a classe é dependente dos atributos, *i.e.*, o conceito embutido nos dados pode ser mapeado por uma função c , tal que $y = c(E)$, onde E é qualquer exemplo do conjunto de dados e y é a classe associada a esse exemplo. Se subconjuntos de atributos diferentes fossem selecionados considerando-se DF1 e DF2, isso significaria que a classe é independente dos atributos, violando assim a conceito do atributo classe.

9 Conclusões

Neste trabalho é proposto o uso da Dimensão Fractal como critério de filtragem de atributos em conjuntos de dados a serem utilizados por algoritmos de aprendizado supervisionado. Vários experimentos, utilizando quatro conjuntos de dados, foram realizados com o objetivo de verificar se a Seleção de um Subconjunto de Atributos, considerando como parâmetro de importância a Dimensão Fractal, melhora ou mantém a precisão do modelo induzido, em relação a outros parâmetros de importância, tais como *Column Importance* e $\mathcal{C}4.5$. Os resultados obtidos usando a DF são animadores. Em termos de precisão, os atributos selecionados permitem, na maioria dos casos, a construção de modelos com performance similar aos outros critérios de importância considerados. Outro resultado que merece destaque é que, considerando-se a percentagem do total de atributos selecionados, o critério de DF escolheu, na média, pouco mais da metade dos atributos definidos como importantes pelo *Column Importance* e pelo $\mathcal{C}4.5$.

É interessante notar que cada conjunto de atributos selecionado é ótimo para o respectivo critério usado como filtro. Por exemplo, o conjunto de atributos selecionado pela Dimensão Fractal apresenta os atributos mais relevantes segundo esse critério. O mesmo acontece para a critério usado por CI e $\mathcal{C}4.5$. Deve ser observado que não existe um consenso sobre qual a melhor medida de importância de um atributo. Essa definição depende da resposta para a pergunta “relevante em relação a que?”. Em geral, a resposta para essa questão está ligada à aplicação para a qual os atributos serão selecionados. Levando em conta o aspecto experimental da área, dado um conjunto de dados, o procedimento mais apropriado é usar diversos métodos e critérios para selecionar, com cada um deles, um subconjunto de

atributos relevantes a fim de verificar com qual (ou quais) desses subconjuntos é possível induzir o melhor classificador.

Um dos trabalhos futuros consiste na análise, com auxílio do especialista do domínio, dos atributos escolhidos pelos diversos critérios. Essa análise permitiria, entre outros, verificar se os atributos selecionados por um método são mais interessantes que os selecionados por outro, do ponto de vista do especialista.

Agradecimentos

Os autores gostariam de agradecer Humberto Razante e Elaine P.M. de Sousa pela valiosa colaboração no desenvolvimento do tema deste trabalho, o qual foi realizado com auxílio parcial da FAPESP.

Referências

- [Blake et al., 1998] Blake, C., Keogh, E., and Merz, C. (1998). UCI repository of machine learning databases. <http://www.ics.uci.edu/mllearn/MLRepository.html>.
- [Faloutsos and Kamel, 1994] 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 (POD94)*, pages 4–13, Minneapolis, MN.
- [Guyon and Elisseeff, 2003] Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning*, 3:1157–1182.
- [John et al., 1994] John, G., Kohavi, R., and Pfleger, K. (1994). Irrelevant features and the subset selection problem. In Kaufmann, M., editor, *Proceedings of the Eleventh International Conference on Machine Learning*, pages 167–173, San Francisco, CA.
- [Kohavi and John, 1997] Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, pages 273–324.
- [Lee, 2000] Lee, H. D. (2000). Seleção e construção de features relevantes para o aprendizado de máquina. Dissertação de Mestrado, ICMC-USP, <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-15032002-113112>.
- [Liu and Motoda, 1998] Liu, H. and Motoda, H. (1998). *Feature Selection for Knowledge and Data Mining*. Kluwer Academic Publishers, Massachusetts.
- [Mandelbrot, 1985] Mandelbrot, B. B. (1985). *The Fractal Geometry of Nature: Updated and Augmented*. W. H. Freeman and Company, New York.
- [Mitchell, 1997] Mitchell, T. M. (1997). *Machine Learning*. WCB McGraw-Hill.
- [Quinlan, 1993] Quinlan, J. R. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann. San Francisco, CA.
- [Schroeder, 1991] Schroeder, M. (1991). *Fractals, Chaos, Power Laws: Minutes from an Infinite Paradise*. W. H. Freeman and Company, New York.
- [Sousa, 2003] Sousa, E. P. M. (2003). Classificação e detecção de agrupamentos usando estruturas de indexação. Monografia apresentada para o Exame de Qualificação de Doutorado, ICMC-USP.
- [Traina et al., 2003] Traina, C., Traina, A. J. M., and Faloutsos, C. (2003). Mde - measure distance exponent manual.
- [Traina et al., 2000] 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.