APLICANDO SELEÇÃO UNILATERAL EM CONJUNTOS DE EXEMPLOS DESBALANCEADOS: RESULTADOS INICIAIS

Gustavo E.A.P.A. Batista[1], André C.P.L.F. Carvalho[2], Maria C. Monard[3]

[1]Institulo de Ciências Matemáticas de São Carlos,
Universidade de São Paulo.
Silicon Graphics Brasil.

[2]Institulo de Ciências Matemáticas de São Carlos,
Universidade de São Paulo.

[3]Institulo de Ciências Matemáticas de São Carlos,
Universidade de São Paulo/ILTC.

{gbatista, andre, mcmonard}@icmc.sc.usp.br

Resumo Em Aprendizado de Máquina supervisionado diversos fatores podem influenciar no desempenho do classificador induzido por um sistema de aprendizado. Entre eles, um fator que pode levar a um desempenho inapropriado e que tem recebido relativamente pouca atenção da comunidade é quando existe uma grande diferença na quantidade de exemplos de cada classe. Neste caso, o sistema de aprendizado pode encontrar dificuldades no aprendizado do conceito referente à classe representada pela minoria dos exemplos. Neste trabalho é discutido um critério para diminuir o número de exemplos da classe majoritária, a fim de melhorar a performance de classificação da classe minoritária. É proposto também o uso da métrica VDM a fim de melhorar o desempenho dos métodos propostos. Por fim, são realizados experimentos com o objetivo de comprovar a eficiência dos métodos para melhorar o desempenho de classificação da classe minoritária.

1 Introdução

Neste artigo é assumido, por simplicidade, que os problemas a serem tratados possuem duas classes. Entretanto, os métodos descritos neste artigo podem ser aplicados a problemas com mais classes. Nesse caso, a ordem em que os métodos de seleção são aplicados a cada uma das classes é relevante.

Para muitos domínios de aplicação, como por exemplo o diagnóstico de doenças raras, é comum que exista uma grande desproporção entre a quantidade de exemplos rotulados em cada uma das classes presentes. Nesses casos, é muito simples projetar um classificador que possua alta precisão, através de um sistema que responda com a classe majoritária para qualquer novo caso. Entretanto, classificar precisamente a classe minoritária é freqüentemente o maior objetivo de tais aplicações.

Muitos dos sistemas de Aprendizado de Máquina (AM) tradicionais não estão preparados para aprender conceitos que classifiquem ambas as classes, com precisão, sob tais condições. Como resultado do aprendizado, freqüentemente obtêm-se uma alta precisão de classificação para a classe majoritária e uma inaceitável precisão para a classe minoritária.

Este trabalho discute métodos para selecionar criteriosamente os exemplos da classe majoritária. Tal seleção criteriosa busca remover os exemplos da classe majoritária que possuem pouca influência no aprendizado do conceito que eles representam, de forma a pouco afetar o aprendizado da classe majoritária. Através de uma seleção dos exemplos da classe majoritária é possível diminuir a desproporção do número de exemplos entre as classes e, desta forma, melhorar o aprendizado da classe minoritária. Tais métodos de seleção de exemplos que buscam diminuir criteriosamente o número de exemplos da classe majoritária são conhecidos por *métodos de seleção unilateral*.

Este trabalho propõe alguns melhoramentos ao método de seleção unilateral proposto em [Kubat 97]. Tal método é baseado em *sistemas de aprendizado baseados em instâncias* (instance-based learning). Sistemas de aprendizado baseados em instâncias aprendem através do armazenamento de exemplos tal que com a utilização de alguma medida de distância, eles são capazes de classificar novos exemplos. Quando todos os atributos são numéricos, a distância Euclidiana pode ser utilizada para comparar os exemplos. Entretanto, na presença de atributos simbólicos, sistemas baseados em instâncias tipicamente utilizam métricas muito simples, tal como contar os atributos que possuem o mesmo valor. Métricas como esta podem falhar na captura da complexidade do domínio do problema, levando a um desempenho inadequado.

2 Problemas com Conjuntos de Exemplos Desbalanceados

Conjuntos de exemplos desbalanceados são de difícil aprendizado, uma vez que a maioria dos sistemas de aprendizado de máquina não está preparado para lidar com uma grande diferença entre o número de exemplos que pertencem a cada classe. Entretanto, problemas reais com essas características são muito comuns em Aprendizado de Máquina, como por exemplo, na recuperação de informação a partir de textos [Lewis 94], no diagnóstico de doenças raras como doenças na tiróide [Blake 98], na detecção de fraudes em operações com cartões de crédito [Stolfo 97], entre outros.

Por que o aprendizado de conceitos, para ambas as classes, é tão complicado sob tais condições? Imagine uma situação como a ilustrada na Figura 1, na qual existe um desbalanço muito grande entre a classe majoritária (-) e a classe minoritária (+), bem como existem exemplos da classe majoritária erroneamente rotulados (ruído). Os exemplos esparsos da classe minoritária (+) podem confundir um classificador como o *k-Nearest Neighbor (k-NN)*. Por exemplo, utilizando 1-NN, muitos exemplos da classe minoritária (+) podem ser classificados erroneamente como sendo da classe majoritária (-), uma vez que o seu vizinho mais próximo é um exemplo da classe majoritária (-) rotulado incorretamente. Em uma situação
extrema, na qual a diferença entre classes é muito grande, a probabilidade do vizinho mais próximo de um caso da classe minoritária (+) ser um caso da classe majoritária (-) se torna alta, e a taxa de erro da classe minoritária pode se aproximar de 100%, o que é inaceitável para muitas aplicações.

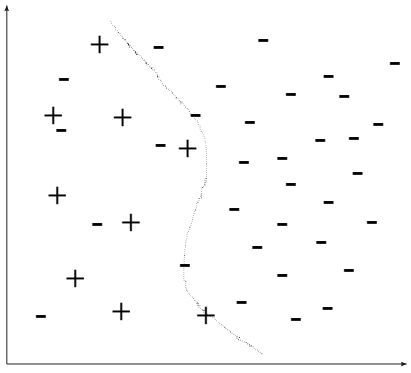

Figura1: Excesso de casos negativos e alguns esparsos casos positivos.

Árvores de decisão também sofrem de um problema similar ao 1-NN. Na presença de ruído, uma árvore de decisão pode se tornar muito especializada (processo conhecido como *overfitting*), ou seja, pode ser necessário criar muitos ramos para distinguir exemplos da classe minoritária (+) dos exemplos da classe majoritária (-), os quais são ruído. No caso de haver poda, o problema pode ainda persistir, uma vez que após a remoção de ramos considerados demasiadamente específicos aos dados, os novos nós folhas são usualmente rotulados com a classe numericamente dominante nesse nó, a qual tem grande probabilidade de ser a classe majoritária do conjunto de exemplos.

Análises teóricas mostram que redes Perceptron Multi-Camadas (MLP) aproximam *probabilidades bayesianas a posteriori* (independentes das probabilidades a priori das classes). Entretanto, análises experimentais têm mostrado que redes neurais MLP possuem dificuldades de aprendizado quando treinadas com conjuntos de exemplos desbalanceados [Barnard 88]. Segundo [Lawrence 98], tal diferença entre análises práticas e teóricas é devido a algumas suposições utilizadas nas análises teóricas, tais como: uma quantidade infinita de dados para treinamento, que o mínimo global seja encontrado, entre outras.

3 Métricas para Medir o Desempenho com Classes Desbalanceadas

A taxa de erro (*E*) ou a precisão (*1-E*) são duas das métricas mais utilizadas para medir a performance de sistemas de aprendizado de máquina. Entretanto, em situações nas quais as probabilidades a priori das classes são muito diferentes, essa medida pode ser enganosa. Por exemplo, em um domínio onde a classe majoritária *x* é representada por 90% dos exemplos, é muito simples obter um classificador com 90% de precisão simplesmente rotulando todo novo exemplo como pertencente a classe *x*. Entretanto, tal procedimento é o mesmo de não realizar aprendizado algum. Outro fator contra considerar somente a precisão (ou taxa de erro) é que esta medida considera diferentes erros cometidos pelo classificador como igualmente importantes. Por exemplo, no diagnóstico médico, diagnosticar um paciente doente como saudável pode ser um erro fatal, enquanto que diagnosticar um paciente saudável como doente é considerado um erro menos sério pois pode ser corrigido com futuros exames. Para problemas onde o custo é relevante, é necessária a definição de uma matriz de custo. Tal matriz define quais são os custos de classificar incorretamente os exemplos segundo os diferentes tipos de erros.

A Tabela 1 mostra uma *matriz de confusão* para um domínio com apenas duas classes, cujo objetivo é discriminar os diferentes tipos de erros e acertos realizados por um classificador.

	Predição Pos.	Predição Neg.
Classe Pos.	Verdadeiro Pos. (*a*)	Falso Neg. (*b*)
Classe Neg.	Falso Pos. (*c*)	Verdadeiro Neg. (*d*)

Tabela 1: Tipos de erros para classificação com duas classes

Além da taxa de erro *(c+b)/(a+b+c+d)*, outras medidas podem ser extraídas da Tabela 1. Dentre elas, duas medem diretamente a taxa de erro de classificação nas classes positiva e negativa:

- *Taxa de Falso Positivo b/(a+b)* é o porcentual de casos erroneamente classificados como pertencentes à classe positiva do total de casos positivos;
- *Taxa de Falso Negativo c/(c+d)* é o porcentual dos casos erroneamente classificados como pertencentes à classe negativa do total de casos negativos;

A Figura 2 [Chan 98] mostra uma relação bastante comum entre a taxa de erro, a taxa de falso positivo e a taxa de falso negativo. Este experimento tem como objetivo identificar transações fraudulentas (classe minoritária, representada como classe negativa) em cartões de crédito, dentre as transações válidas (classe majoritária, representada como classe positiva). Chan e Stolfo treinaram o sistema de aprendizado C4.5 [Quinlan 88] com diferentes distribuições de classes no conjunto de treinamento. O gráfico inicia representando um conjunto de treinamento com 10% de exemplos da classe minoritária e 90% de classe majoritária, e a proporção de exemplos da classe minoritária no conjunto de treinamento é aumentada em 10% a cada iteração. Conforme mostra o gráfico, o aumento de exemplos da classe minoritária faz com que os exemplos dessa classe sejam melhor classificados pelo sistema, entretanto a performance de classificação da classe majoritária é comprometida. A taxa de erro cresce influenciada no mau desempenho da classe majoritária, uma vez que esta classe domina o conjunto de teste.

Esse experimento mostra uma perda de precisão de classificação da classe majoritária com o aumento do número de casos da classe minoritária. Em situações reais, nas quais os exemplos da classe minoritária não podem ser facilmente obtidos, resta a opção de remover os exemplos da classe majoritária na tentativa de atingir uma melhor proporção entre as classes. Entretanto, será possível remover exemplos da classe majoritária de forma a diminuir a perda de precisão de classificação dessa classe? Esse é o principal objetivo da seleção unilateral.

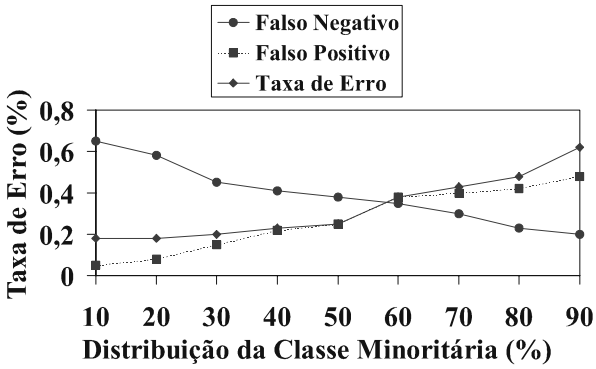

Figura 2: Taxas de erro com diversas distribuições de classes no conjunto de treinamento.

4 Seleção Unilateral

Em AM, o problema do desbalanceamento de classes tem sido estudado e algumas soluções têm sido propostas. Uma delas, a seleção unilateral [Kubat 97] propõe a remoção criteriosa de exemplos pertencentes à classe majoritária, e a manutenção dos exemplos da classe minoritária, uma vez que tais exemplos são muito raros para serem perdidos, ainda que alguns deles sejam ruído. Tal remoção criteriosa consiste em detectar e remover, através de alguma heurística, exemplos que são menos confiáveis. Essa heurística pode ser melhor entendida se os exemplos forem divididos em quatro grupos distintos:

1. Exemplos rotulados erroneamente (ruído). Por exemplo, os exemplos da classe majoritária (-) que estão na parte esquerda da região de decisão da Figura 1;
2. Exemplos redundantes. Tais exemplos podem ser representados por outros exemplos que já estão no conjunto. Por exemplo, os exemplos que estão longe da borda de decisão, tais como os localizados na parte superior direita da Figura 1;
3. Exemplos próximos à borda de decisão. Esses exemplos são pouco confiáveis pois mesmo uma pequena quantidade de ruído em alguns valores de seus atributos pode colocá-los na região de decisão errada;
4. Exemplos seguros. São aqueles que não estão muito próximos da borda de decisão nem muito distantes dela, esses exemplos devem ser mantidos para o aprendizado.

A técnica de seleção unilateral busca criar um conjunto de treinamento que consista somente de exemplos seguros. Desta forma, exemplos da classe majoritária incorretamente rotulados, exemplos muito próximos à borda de decisão e exemplos redundantes devem ser eliminados.

Exemplos situados na borda de decisão e exemplos incorretamente rotulados podem ser detectados através

do conceito de ligações Tomek [Tomek 76], definido a seguir. Dados dois exemplos *x* e *y*, de forma que cada um pertença a uma classe diferente, seja a distância entre *x* e *y* denotada por *d(x, y)*, então o par *(x, y)* é chamado de ligação Tomek caso não exista um exemplo *z* tal que *d(x, z) < d(x, y)* ou *d(y, z) < d(y, x)*. Exemplos que participam de ligações Tomek ou estão próximos à borda de decisão ou são ruído. A Figura 3 ilustra o novo conjunto de exemplos obtido após a remoção de exemplos da classe majoritária que formam ligações Tomek no conjunto original de exemplos (Figura 1).

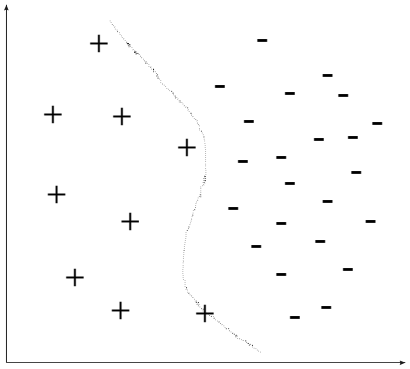

Figura 3: Conjunto de exemplos sem ruído e casos próximos à borda de decisão.

Para remover exemplos redundantes pode-se criar um subconjunto consistente, *C*, do conjunto original de exemplos, *S*. Por definição, um conjunto $C \subseteq S$ é consistente com *S* se, utilizando-se o classificador *k*-nearest neighbor com *k=1* (1-NN), ele corretamente classifica os exemplos de *S*. Como o objetivo não é encontrar o menor subconjunto consistente possível, mas apenas fazer com que o número de exemplos da classe majoritária diminua, foi proposta em [Kubat 97] uma variação da técnica original criada em [Hart 68] para encontrar subconjuntos consistes. A idéia é selecionar aleatoriamente um exemplo da classe majoritária juntamente com todos os exemplos da classe minoritária e movê-los para um conjunto *C*. Então, utiliza-se uma regra 1-NN com os exemplos em *C* para re-classificar os exemplos em *S*. Todos os exemplos de *S* que forem incorretamente classificados são movidos para *C*. A Figura 4 ilustra o novo conjunto de exemplos obtido com a remoção de exemplos redundantes da classe majoritária do conjunto original (Figura 1).

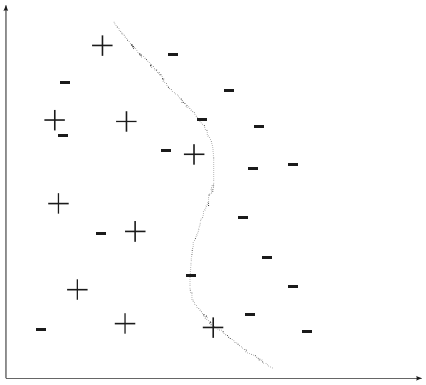

Figura 4: Conjunto de exemplos sem casos redundantes.

5 A Métrica VDM para Calcular Distância entre Atributos Simbólicos

Em [Stanfill 86] foi apresentado um poderoso método para medir a distância entre atributos simbólicos, chamado *Value Difference Metric (VDM)*. Ao contrário dos métodos mais simples que medem a distância entre atributos simbólicos simplesmente contando o número de atributos com o mesmo valor, o método VDM considera a similaridade de classificação para cada possível valor de cada atributo. Através deste método é criada uma matriz, baseada nos exemplos do conjunto de treinamento, contendo a distância entre os diversos valores de cada atributo. A distância *d* entre dois valores para um determinado atributo *V* é:

$$d(V_1, V_2) = \sum_{i=1}^{n} \left| \frac{C_{1_i}}{C_1} - \frac{C_{2_1}}{C_2} \right|^k$$

Nesta equação, V_1 e V_2 são dois possíveis valores para o atributo *V*. A distância entre os valores é a soma sobre todas as *n* classes. C_{1i} é o número de exemplos nos quais o valor V_1 do atributo *V* ocorreu juntamente com a classe *i*, C_1 é o número total de exemplos nos quais o valor V_1 ocorreu nos exemplos, e *k* é uma constante, freqüentemente 1.

A métrica VDM estabelece que dois valores são similares se eles ocorrem com a mesma freqüência relativa para todas as classes. O termo C_{1i}/C_1 representa a probabilidade do exemplo ser classificado com a classe *i* dado que o atributo em questão possui valor V_1. Desta forma, a similaridade entre dois valores é calculada encontrando-se a soma das diferenças das probabilidades sobre todas as classes.

Considere o seguinte exemplo [Cost 93]. Em um conjunto de exemplos, um determinado atributo pode assumir três valores *A*, *B* e *C*, e duas classes são possíveis *x* e *y*. A partir dos dados é possível construir

uma tabela como mostra a Tabela 2, a qual representa o número de vezes que o atributo em questão assumiu cada um dos valores possíveis para as duas classes. A partir da Tabela 2, e utilizando a métrica VDM, é possível construir a Tabela 3 que representa a distância entre cada um dos valores que o atributo pode assumir. A freqüência de ocorrência de *A* para a classe *x* é de 57.1%, uma vez que existem 4 exemplos classificados como *x* em um total de 7 exemplos com o valor *A*. De forma similar, as freqüências de ocorrência para *B* e *C* são 28.6% e 66.7%, respectivamente. A freqüência de ocorrência de *A* para a classe *y* é de 42.9% e assim por diante. Por exemplo, a distância entre *A* e *B*, através da métrica VDM, resulta em $|4/7 - 2/7| + |3/7 - 5/7| = 0.571$. A Tabela 3 mostra todas as distâncias calculadas.

	Classes	
Valores do atributo	*x*	*y*
A	4	3
B	2	5
C	4	2

Tabela 2: Total de ocorrências de cada valor para cada classe.

	Valores do Atributo		
	A	B	C
A	0.000	0.571	0.191
B	0.571	0.000	0.762
C	0.191	0.762	0.000

Tabela 3: Distâncias entre valores segundo a métrica VDM.

A métrica VDM possui características de medida de distância, são elas:

1. $d(a, b) > 0, a \neq b$;
2. $d(a, b) = d(b, a)$;
3. $d(a, a) = 0$;
4. $d(a, b) + d(b, c) \geq d(a, c)$.

6 Experimentos Investigativos Iniciais

Para realizar uma investigação inicial do funcionamento dos métodos de seleção unilateral, foi selecionado o conjunto de exemplos *Breast Cancer* do repositório UCI [Blake 98]. Com esse conjunto é possível visualizar em três dimensões dois grupos de exemplos que representam as classes *cancer=benign* e *cancer=malignant*, além de alguns exemplos esparsamente distribuídos. Os eixos utilizados na visualização são constituídos pelos seguintes atributos: *clump thickness*, *uniformity of cell size* e *bare nuclei*. A Figura 5 mostra o conjunto de dados original visualizados através da ferramenta MineSet.

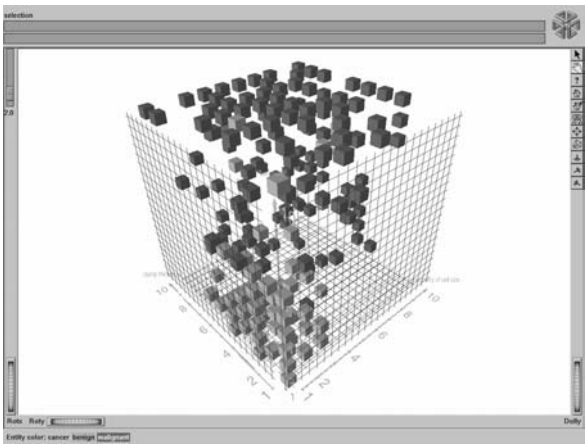

Figura 5 Conjunto de Exemplos Breast Cancer Original.

Para uma melhor visualização, foram removidos os exemplos que constituíam ligações Tomek da classe *cancer=benign* (representada em cor azul), uma vez que esta classe apresenta alguns exemplos esparsamente distribuídos, os quais possivelmente podem ser ruído. E foi aplicado subconjuntos consistentes nos exemplos da classe *cancer=malignant* (representada em cor vermelha), uma vez que esta classe possui muitos exemplos distantes da borda de decisão. A Figura 6 mostra o conjunto de exemplos após a aplicação dos métodos de seleção unilateral.

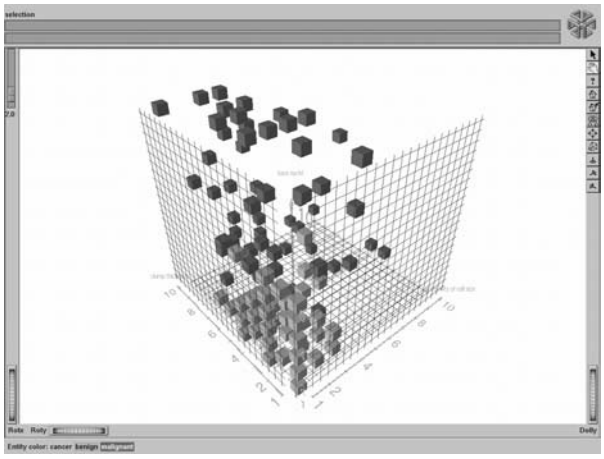

Figura 6: Conjunto de Exemplos Breast Cancer sem Exemplos Redundantes, Próximos à Borda e Ruído.

7 Resultados Experimentais

Foram realizados diversos experimentos para verificar se, para aplicações com conjuntos de exemplos desbalanceados, a seleção unilateral pode melhorar o desempenho de sistemas de aprendizado sobretudo para a classificação de casos da classe minoritária. O sistema de aprendizado selecionado para este experimento foi o C5.0 [Quinlan 88]. Foi escolhido

também o conjunto de exemplos *Hepatitis* proveniente do repositório UCI [Blake 98]. O conjunto hepatitis possui 155 exemplos com 123 (79,3%) pertencentes à classe *live* e 32 casos (20,6%) pertencentes à classe *die*. Durante o experimento foram utilizadas as facilidades do ambiente AMPSAM [Batista 97] para facilitar o processo de medição do desempenho de classificação.

O conjunto Hepatitis é conhecido na comunidade de Aprendizado de Máquina por ser difícil de obter resultados satisfatórios de classificação. Segundo [Holte 93] muitos poucos sistemas de aprendizado obtiveram uma precisão superior a dois pontos percentuais acima da *baseline accuracy*, ou seja, precisão dois pontos percentuais acima de 79,3%.

O desempenho do algoritmo C5.0 foi medido sobre o conjunto original com todos os exemplos (a); sobre o conjunto original retirando-se ruído e exemplos de borda através de ligações Tomek (b); sobre o conjunto original retirando-se exemplos redundantes através de 1-NN (c); sobre o conjunto original retirando-se ruído, exemplos de borda e exemplos redundantes (d); e finalmente sobre um conjunto de exemplos no qual exemplos da classe majoritária foram removidos aleatoriamente (e). As taxas de erro foram obtidas através do método de resampling 3-fold cross-validation. O número de partições $k=3$ foi escolhido uma vez que existem muitos poucos exemplos da classe minoritária, e maiores valores de k podem levar a resultados com alta variância. Para confirmar os resultados, o experimento foi repetido três vezes. Os experimentos mostraram resultados semelhantes, por motivos de espaço apenas o resultado de um dos experimentos é mostrado na Tabela 4. A coluna *N* mostra a proporção de exemplos entre as classes no conjunto de treinamento. *Fn* e *Fp* são as taxas de falso negativo e falso positivo, respectivamente. As colunas $\sigma(Fn)$ e $\sigma(Fp)$ mostram os desvios padrões para as taxas de falso negativo e falso positivo.

	N	Fn	σ(Fn)	Fp	σ(Fp)
a	80/23	10%	5,8%	62%	5,1%
b	70/23	15%	4,0%	43%	11,7%
c	62/23	13%	8,6%	57%	11,7%
d	55/23	28%	4,5%	29%	7,9%
e	50/23	9%	2,6%	57%	17,4%

Tabela 4: Taxas de Erro e Desvios padrão.

Os resultados sugerem que a seleção unilateral pode abaixar a taxa de erro da classe *die* (minoritária) consideravelmente, em especial quando ligações Tomek são utilizadas (*b* e *d*). Entretanto, a seleção aleatória (a qual não utiliza heurística alguma) obteve resultados comparáveis à seleção por sub-conjuntos consistentes. Apesar de que a seleção aleatória não utilizar heurística alguma, ela possui o mérito de remover os exemplos com igual probabilidade. Desta forma, dentre os métodos de seleção utilizados, a remoção aleatória é aquela que provavelmente realiza a menor mudança na distribuição original dos exemplos. Podem existir, ainda, outras razões para tal efeito, entre elas: é sabido que os conjuntos de exemplos do repositório UCI já foram largamente analisados mesmo antes de serem disponibilizados no repositório. Durante tais análises muitos dos exemplos de borda, ruído e redundantes podem ter sido retirados, fazendo com que a eficácia das heurísticas fosse reduzida. Um outro motivo é a não remoção de qualquer exemplo da classe minoritária, mesmo que possivelmente ruído. Tal atitude foi tomada pela necessidade de se manter os poucos exemplos da classe minoritária existentes. Entretanto, o ruído presente nos exemplos da classe minoritária pode diminuir a precisão de classificação. Atualmente são retirados somente exemplos da classe majoritária que formam ligações Tomek, entretanto, tais ligações podem estar relacionadas a exemplos erroneamente rotulados da classe minoritária. Retirar somente exemplos da classe majoritária que compõem ligações Tomek não somente não retira exemplos com ruído da classe minoritária, como também pode retirar exemplos da classe majoritária importantes, os quais não são ruído nem próximos à borda de decisão. Uma vez que exemplos da classe minoritária são preciosos, uma provável melhoria seria diferenciar entre exemplos que são ruído e exemplos que estão na borda de decisão. Infelizmente, ligações Tomek não fornecem um meio seguro de diferenciar tais tipos de exemplos e outros métodos necessitam ser investigados.

8 Conclusões

Uma grande diferença entre a quantidade de exemplos pertencentes a cada uma das classes é uma característica comum nos problemas de Aprendizado de Máquina. Nesse tipo de problema, pode ser trivial criar um sistema que classifique tais exemplos com muito boa precisão, entretanto um problema muito mais complexo é obter uma boa classificação para a classe minoritária. Seleção unilateral busca retirar, criteriosamente, exemplos da classe majoritária de forma a aproximar o número de exemplos pertencentes à
cada classe. Este trabalho mostra alguns resultados que sugerem a eficácia dessa técnica. Como trabalhos futuros ira-se procurar por heurísticas que possam distinguir entre exemplos com ruído e exemplos próximos à borda de decisão, bem como novas técnicas para selecionar exemplos da classe majoritária.

Agradecimentos

Trabalho realizado com auxilio parcial da FINEP e da Silicon Graphics Brasil.

Referências Bibliográficas

[Barnard 88] Barnard, E.; Cole, R.A.; Hou, L. *Location and Classification of Plosive Constants Using Expert Knowledge and Neural Nets Classifiers.* Journal of the Acoustical Society of America, 1988, 84 Supp 1:S60.

[Batista 97] Batista, G.E.A.P.A.; Monard, M.C. *AMPSAM: Um Ambiente Computacional para Medir Performance de Sistemas de Aprendizado de Máquina.* Anais do I ENIA, 1997.

[Blake 98] Blake, C.; Keogh, E.; Merz, C.J. *UCI Repository of Machine Learning Databases.* Irvine, CA: University of California, Department of Information and Computer Science. http://www.ics.uci.edu/~mlearn/MLRepository.html.

[Chan 98] Chan, P.K.; Stolfo, S.J. *Learning with Non-uniform Class and Cost Distributions: Effects and a Distributed Multi-Classifier Approach.* KDD-98 Workshop on Distributed Data Mining, 1998, p~1-9.

[Cost 93] Cost, S.; Salzberg, S. *A Weighted Nearest Neighbor Algorithm for Learning with Symbolic Features.* Machine Learning, 10 (1), 1993, pp. 57-78.

[Hart 68] Hart, P.E. *The Condensed Nearest Neighbor Rule.* IEEE Transactions on Information Theory, IT-14, 1968, pp. 515-516.

[Holte 93] Holte, C.R. *Very Simple Classification Rules Perform Well on Most Commonly Used Datasets.* Machine Learning, 11, 1993, pp.~63-91.

[Kubat 97] Kubat, M.; Matwin, S. *Addressing the Course of Imbalanced Training Sets: One-Sided Selection.* Draft, Department of Computer Science, University of Ottawa, 1997.

[Lawrence98] Lawrence, S.; Burns, I.; Back, A.; Tsoi, A.C.; Giles, C.L. *Neural Network Classification and Prior Class Probabilities.* Tricks of the trade, Lecture Notes in Computer Science State-of-the-art surveys, G. Orr, K.R. Müller, R. Caruana (editors), Springer Verlag, 1998, pp.~299-314.

[Lewis 94] Lewis, D.; Catlett, J. *Heterogeneous Uncertainty Sampling for Supervised Learning.* Proceedings of the 11th International Conference on Machine Learning, ICML94, Morgan Kaufmann, 1994, pp. 148-156.

[Quinlan 88] Quinlan, J.R. *C4.5 Programs for Machine Learning.* Morgan Kaufmann Publishers, CA, 1988.

[Stanfill 86] Stanfill, C.; Waltz, D. *Toward Memory-Based Reasoning.* Communications of the ACM, 29(12), 1986, pp. 1213-1228.

[Stolfo 97] Stolfo, S.J.; Fan, D.W.; Lee, W.; Prodromidis, A.L.; Chan, P.K. *Credit Card Fraud Detection Using Meta-Learning: Issues and Initial Results.* Proc. AAAI-97 Workshop on AI Methods in Fraud and Risk Management, 1997.

[Tomek 76] Tomek, I. *Two Modifications of CNN.* IEEE Transactions on Systems Man and Communications, SMC-6, 1976, pp. 769-772.