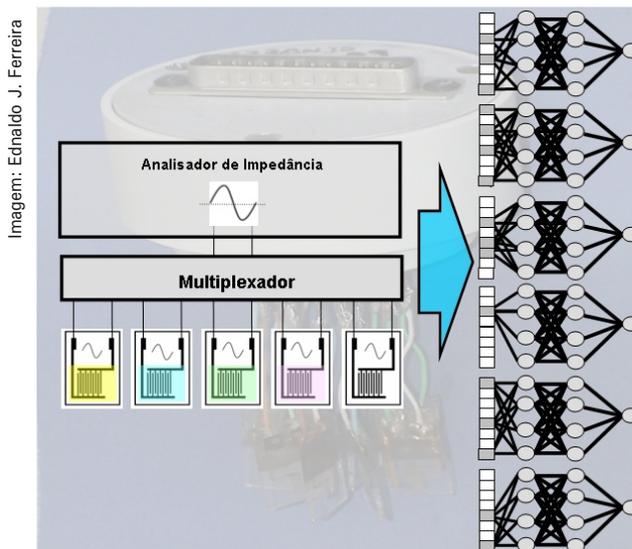


Algoritmo Genético para Construção de Ensembles de Redes Neurais: Aplicação à Língua Eletrônica

Introdução



Aprendizado de máquina (AM) é uma área da inteligência artificial (IA) cujo objetivo é o desenvolvimento de algoritmos capazes de adquirir conhecimento de forma automática. Em geral, os algoritmos de AM induzem um conceito a partir de um conjunto de dados de exemplo. A indução é uma operação que consiste em se estabelecer uma verdade universal ou uma proposição geral com base em um certo número de dados singulares ou proposições de menor generalidade. O algoritmo de indução (indutor) ou algoritmo de aprendizado constrói uma hipótese do domínio com base em um conjunto de exemplos. Em aprendizado supervisionado, um exemplo é descrito por um conjunto de atributos

(características), geralmente de domínios diferentes, acompanhado por um rótulo associado. Em dados passíveis de classificação, o rótulo corresponde à classe do exemplo.

Os métodos de aprendizagem indutiva funcionam melhor quando são alimentados com características relevantes (RUSSELL; NORVIG, 2004). Características irrelevantes ou redundantes podem confundir o algoritmo de aprendizado, ajudando a esconder as distribuições de pequenos conjuntos de características realmente relevantes (KOLLER; SAHAMI, 1996). Por exemplo, na análise de uma doença hematológica, o código de endereçamento postal, o número do telefone e o salário do paciente são características totalmente irrelevantes para o diagnóstico e que podem prejudicar a qualidade de uma hipótese induzida. A propensão à excessiva quantidade, redundância e irrelevância de atributos é também uma realidade dos conjuntos de dados provenientes de uma língua eletrônica.

A língua eletrônica (LE) é um dispositivo eletroquímico composto por um conjunto de unidades sensoriais, diferenciadas por polímeros condutores, que captam algumas propriedades químicas da substância em análise, propiciando assim a identificação do sabor. O sistema de aquisição de dados da LE é caracterizado por grande permissividade na variação dos parâmetros. Esses parâmetros incluem as frequências de operação dos sensores, a combinação de tipos de polímeros utilizados na construção do sensor, a amplitude do sinal aplicado, entre outros. Como são desconhecidas as melhores configurações dos parâmetros, o que se faz, na prática, é coletar dados em uma ampla variação paramétrica. Quanto maior a variação dos parâmetros considerada na coleta de cada amostra em análise, maior a quantidade de atributos descritores. Embora a quantidade de atributos seja, em princípio, uma qualidade do sistema de aquisição, para alguns algoritmos de aprendizado ela pode representar um problema conhecido como maldição da dimensionalidade (KOLLER; SAHAMI, 1996), o que provoca a indução de hipóteses com baixa acurácia. Experimentos realizados com classificadores simples, de redes neurais artificiais (RNAs), para categorização qualitativa de amostras de cafés, não demonstraram desempenhos satisfatórios¹. Isso é decorrente, principalmente, da grande quantidade de atributos irrelevantes e redundantes que descrevem os dados.

¹Resultados insatisfatórios do treinamento de redes neurais com dados da LE para distinção da qualidade de cafés torrados, diluídos em água, podem ser observados na seção de resultados.

Autores

Ednaldo José Ferreira
Ciência da Computação, MSc.
Embrapa Instrumentação
Agropecuária, C.P. 741,
CEP 13560-970, São Carlos, SP
ednaldo@cnpdia.embrapa.br

Alexandre Claudio Botazzo Delbem
Engenheiro Elétrico, Dr.
ICMC - USP
CEP 13560-970, São Carlos, SP
acbd@icmc.usp.br

Os classificadores induzidos a partir dos conjuntos de dados da LE e de muitos outros conjuntos de diversos domínios estão sujeitos à baixa acurácia, devido à demasiada presença de atributos irrelevantes e/ou redundantes. Para minimizar esse problema e melhorar a acurácia de um classificador, duas abordagens podem ser utilizadas (BERNARDINI, 2002) seleção de atributos e *ensembles* de classificadores.

A seleção de atributos relevantes (FSS *Feature Subset Selection*) é um dos temas pesquisados em AM. Embora a maioria dos algoritmos de AM busque, de alguma forma, selecionar ou atribuir graus de importância aos atributos dos dados, estudos indicam que a presença de muitos atributos irrelevantes deteriora o desempenho desses algoritmos na construção das hipóteses. Por isso, métodos explícitos de seleção de atributos podem aumentar a acurácia dos classificadores induzidos. No contexto da LE, o processo de seleção de atributos relevantes representa mais que um meio para aumentar o desempenho do algoritmo de aprendizado, pois a redução do conjunto de atributos pode também:

- otimizar o tempo de aquisição dos dados: quanto menor o número de parâmetros considerados na coleta, menor o tempo de aquisição dos dados;
- reduzir custos com a construção do arranjo de sensores: quanto menor o número de sensores, menor o custo envolvido na construção de uma LE;
- indicar as melhores freqüências de operação e os melhores polímeros condutores: a definição da melhor freqüência e tipo de polímero condutor aumenta a sensibilidade e seletividade da LE;
- viabilizar a construção de equipamentos portáteis e simplificados: um conjunto reduzido de freqüências viabiliza a construção de equipamentos portáteis dotados de uma eletrônica relativamente simples.

A segunda abordagem é a combinação de classificadores diversos em um comitê: *ensemble*. Um *ensemble* é formado por um conjunto de classificadores cujas decisões individuais são combinadas de alguma forma para classificar um novo caso. Os *ensembles* são freqüentemente mais acurados que qualquer um dos seus classificadores componentes (DIETTERICH, 2000)

Uma forma efetiva para construção de bons *ensembles* é por meio da manipulação do conjunto de atributos. A tarefa de encontrar subconjuntos de atributos para construção de um *ensemble* de classificadores acurados e diversos é conhecida como seleção de características para *ensembles* (EFS - *Ensemble Feature Selection*) (OPITZ, 1999). O processo de EFS é caracterizado pela junção das duas abordagens supracitadas. Enquanto os algoritmos tradicionais de seleção de atributos têm por objetivo encontrar o melhor subconjunto de atributos relevantes tanto para tarefa de aprendizado quanto para o algoritmo de AM, EFS tem o objetivo adicional de encontrar subconjuntos que promovam diversidade de predição entre os classificadores (OPITZ, 1999; TSYMBAL et al., 2003a) Dentre os diversos algoritmos para EFS, destacam-se os algoritmos genéticos (OPITZ,

1999a; OPITZ; SHAVLIK, 1999; TSYMBAL et al., 2003a; TSYMBAL et al., 2003b).

O emprego de algoritmo genético (AG) para EFS foi proposto por Opitz (1999) para um *ensemble* de classificadores de RNAs. O algoritmo GEFS (*Genetic Ensemble Feature Selection*) mostrou resultados comparáveis e até superiores aos métodos tradicionais para construção de *ensembles*, como *Bagging* (BREIMAN 1996) e *AdaBoost* (FREUND; SCHAPIRE, 1996; HAYKIN, 2001).

Embora GEFS seja eficaz para construção de um *ensemble* de classificadores de RNAs, ele não tem o compromisso de utilizar um subconjunto mínimo de atributos para construção do classificador final (*ensemble*). Ou seja, GEFS está focado na seleção de subconjuntos de atributos para aumento da acurácia e da diversidade entre os classificadores, mas isso não provê garantias de que, ao final da execução, um subconjunto reduzido dos atributos totais seja resultante do processo. Esta peculiaridade é desinteressante para bases de dados com alta redundância e irrelevância, como são àquelas provenientes de uma LE. A falta de compromisso com a redução do total de atributos impede que GEFS possa usufruir os méritos atribuídos a um método de FSS. Mesmo utilizando intrinsecamente o processo de seleção de atributos para gerar os classificadores, não se pode afirmar que GEFS sempre se comportará globalmente como um método de FSS. Mais especificamente, para os dados da LE isso significa que todos os atributos podem ser utilizados na construção do *ensemble* inviabilizando a determinação das melhores freqüências de operação, dos tipos de polímeros condutores e de todos os outros parâmetros considerados na coleta de dados.

Ferreira e Delbem (2006b) modificaram o algoritmo GEFS original para conjugar o processo de EFS e de FSS no mesmo algoritmo. O algoritmo, denominado GEFS Modificado, mostrou-se eficaz na construção de *ensembles* de RNAs assim como na redução do número de atributos dos dados de diversos domínios. Essa característica de GEFS Modificado torna-o promissor para aplicações com a LE e por isso vem sendo adotado como o algoritmo ideal para construção de classificadores. Nesse contexto, esta circular tem como objetivo mostrar a potencialidade de GEFS Modificado e sua metodologia de validação e certificação, assim como o atual estágio de pesquisa e aplicação com os dados sobre a qualidade global de cafés coletados com uma LE.

Seleção de atributos para *ensembles*

A amostragem de exemplos de treinamento realizada por métodos como *Bagging* (BREIMAN, 1996) e *AdaBoost* (FREUND; SCHAPIRE, 1996; HAYKIN, 2001) é um mecanismo comprovadamente eficiente para obtenção de classificadores acurados e diversos. Para esses métodos, a diversidade preditiva² é consequência das diferentes amostras, retiradas do conjunto de dados, utilizadas no treinamento dos classificadores. Outra forma efetiva para se obter diversidade é por meio do arranjo de diferentes

²A diversidade preditiva é um requisito para a construção de *ensembles* de classificadores adequados. A medida de diversidade de um classificador reflete o seu desacordo ou divergência de predição de classes em relação aos demais classificadores do *ensemble*.

subconjuntos de atributos. A tarefa de encontrar subconjuntos de atributos adequados para a construção de um *ensemble* de classificadores acurados e diversos é conhecida como EFS (OPITZ, 1999).

Enquanto os algoritmos tradicionais de FSS têm por objetivo encontrar o melhor subconjunto de atributos, relevante tanto para tarefa de aprendizado quanto para o algoritmo de AM, EFS tem o objetivo adicional de encontrar subconjuntos que promovam desacordo (diversidade preditiva) entre os classificadores (OPITZ, 1999; TSYMBAL et al., 2003a). Assim como nas técnicas tradicionais de FSS, EFS se posiciona em relação a quatro pontos fundamentais:

- ponto de partida (condição inicial do *ensemble*);
- critério de parada;
- estratégia de avaliação ou função de avaliação (*fitness*);
- estratégia ou organização da busca.

O ponto de partida estabelece quais são os subconjuntos iniciais de atributos que cada um dos classificadores utiliza como referência para ser construído e inferir sua predição. Geralmente, uma seleção aleatória de subconjuntos é adotada como ponto de partida. A simples seleção randômica (RS - *Random Subspacing*) estabelecida como condição inicial pode ser uma técnica efetiva para construção de um *ensemble* (HO, 1998).

A função de avaliação ou função de aptidão (*fitness function*) tem o papel de representar numericamente o compromisso entre a acurácia e a diversidade preditiva do classificador construído para um dado subconjunto de atributos. A função de avaliação da Equação 1, proposta por Opitz (1999), é uma das mais adotadas em EFS.

$$Fitness_i = Acc_i + \lambda Div_i \quad (1)$$

onde Acc_i é a acurácia e Div_i a diversidade preditiva do i -ésimo classificador; λ é o coeficiente que define o peso da diversidade.

A busca por subconjuntos de atributos que proporcionem a construção de classificadores acurados e diversos é dependente da estratégia de busca adotada. Tsybmal et al. (2003a), compararam quatro importantes estratégias de busca para EFS: *hill-climbing* (1), Seleção Seqüencial Adiante (2), Seleção Seqüencial para Trás (3) e Algoritmo Genético (4). Eles analisaram as quatro estratégias de busca para cinco diferentes maneiras de quantificar a diversidade preditiva em 21 bases de dados. O algoritmo genético se apresentou como a melhor estratégia para todas as medidas de diversidade.

A utilização de algoritmo genético como estratégia para construção de *ensembles* de RNAs foi inicialmente proposta por Opitz e Shavlik (1999). O ADDEMUP (*Accurate and Diverse Ensemble-Maker giving United Prediction*), como foi denominado o algoritmo, usava AGs para EFS, mas era muito complexo, limitado e não trabalhava apropriadamente com alguns tipos de problema. Inspirado no ADDEMUP, Opitz (1999) propõe um algoritmo, denominado GEFS, com melhorias consideráveis e adequadas para aplicação em uma abordagem *wrapper*. O algoritmo GEFS usa o poder de otimização do AG para melhorar o *ensemble*, gradativamente, com a evolução da população.

Genetic Ensemble Feature Selection

A aplicação de AGs como estratégia de busca tem evidenciado um nicho promissor para pesquisas de FSS e EFS. A primeira proposição eficaz para o uso de AG para EFS foi realizada por Opitz (1999) para construção de um *ensemble* de classificadores de RNAs do tipo MLP (*Multi-Layer Perceptron*) treinadas por *back-propagation* (RUMELHART et al., 1986). O AG para EFS proposto GEFS inicia com uma população de classificadores treinados a partir de subconjuntos distintos de atributos amostrados aleatoriamente, como em RS (HO, 1998). Então, GEFS utiliza operadores genéticos de cruzamento e mutação para gerar novos indivíduos. GEFS treina esses novos indivíduos e mede a acurácia e a diversidade preditiva de cada um deles. Após uma normalização, uma medida da aptidão de cada indivíduo é realizada por meio da Equação 1. Os melhores indivíduos, representados pela sua respectiva rede, são mantidos e os piores descartados. Finalizando o processo, o coeficiente λ é ajustado como forma de equilibrar o peso da diversidade e dar suporte para melhorias nas próximas gerações. O Algoritmo 1 detalha os passos de GEFS para construção de um *ensemble* de RNAs.

Algoritmo 1: Algoritmo GEFS

Gerar População Inicial de N indivíduos usando RS;
 Treinar RNAs a partir dos indivíduos da População Inicial;
Enquanto critério de parada Insatisfeito **faça**
 Aplicar operador de Cruzamento;
 Aplicar operador de Mutação;
 Treinar redes com os novos indivíduos;
 Medir a acurácia;
 Medir a diversidade preditiva com respeito à população corrente;
 Normalizar acurácia e diversidade preditiva;
 Calcular a aptidão de cada Indivíduo ($Fitness_i = Acc_i + \lambda Div_i$);
 Selecionar os N indivíduos mais aptos;
 Ajustar λ ;
 Definir a população como o *ensemble* de RNAs corrente.

Algoritmo GEFS Modificado

Se ao final de um processo de EFS um subconjunto de atributos não pertencer ao conjunto dos atributos utilizados pelo *ensemble*, então esse subconjunto é dispensável para a tarefa de classificação. Essa característica, que é peculiar às técnicas de FSS, permite redução e melhor compreensão dos dados, principalmente nas bases com atributos altamente redundantes e irrelevantes com no caso da LE.

O mecanismo responsável pelo controle da demanda por diversidade preditiva da população em GEFS original é fundamentalmente determinado pelo coeficiente de diversidade lambda (λ) na função de aptidão. A dinâmica de ajuste de λ durante a evolução de GEFS funciona como um sistema de controle automático. Entre duas gerações consecutivas o erro cometido pelo *ensemble* pode sofrer variações (aumentar ou diminuir). Quando esse erro aumenta duas situações são possíveis:

- o erro médio da população não está aumentando e a diversidade média da população está diminuindo;
- o erro médio da população está aumentando e a diversidade média não está diminuindo.

Na primeira situação o coeficiente λ é aumentado (10% do valor corrente) como forma de corrigir a perda de diversidade preditiva da população atual que compõe o *ensemble*. Na segunda situação, inversa à primeira, o coeficiente λ é diminuído (10% do valor corrente) como forma de reduzir o impacto da diversidade preditiva da população atual para a geração seguinte.

O ajuste de λ viabiliza condições para que indivíduos mais diversos ou menos diversos façam parte da população nas gerações seqüentes. Entretanto, quando os indivíduos da população atual são semelhantes, ou seja, há pouca diversidade populacional, outra forma de diversificação é requerida para obtenção de diversidade preditiva. GEFS originalmente utiliza uma alta taxa de mutação para provocar diversidade genética e assim prover indivíduos com maior diversidade preditiva. No entanto, esse artifício minimiza significativamente a probabilidade de um atributo ser descartado durante o processo de construção do *ensemble*. Para garantir que a taxa de mutação corrobore tanto a diversidade preditiva quanto a redução de características totais (RCT), Ferreira e Delbem (2006b) modificaram GEFS (GEFS Modificado), inserindo uma função de adaptação dinâmica para a taxa de mutação. Essa função trabalha em sincronia com a função de ajuste de λ . O ajuste da taxa de mutação é regido pela seguinte equação:

$$p_{g+1} = p_g \omega \frac{(\lambda_{g+1} - \lambda_g)}{|\lambda_{g+1} - \lambda_g|} \quad (2)$$

Para $\omega = 1$

onde p_g é a taxa de mutação da geração g , λ_g é o coeficiente de diversidade para geração g e ω é o fator multiplicador de p_g .

No Algoritmo 2 é descrito o algoritmo GEFS Modificado.

Algoritmo 2: Algoritmo GEFS Modificado

Gerar População Inicial de N indivíduos usando RS;
 Treinar RNAs a partir dos indivíduos da População Inicial;
Enquanto critério de parada Insatisfeito **faça**
 Aplicar operador de Cruzamento;
 Aplicar operador de Mutação;
 Treinar redes com os novos indivíduos;
 Medir a acurácia;
 Medir a diversidade preditiva com respeito à população corrente;
 Normalizar acurácia e diversidade preditiva;
 Calcular a aptidão de cada Indivíduo ($Fitness_i = Acc_i + \lambda Div_i$)
 Selecionar os N indivíduos mais aptos;
 Ajustar λ ;
 Ajustar p ; ***** MODIFICAÇÃO *****
 Definir a população corrente como o *ensemble* de RNAs.

O ajuste faz a taxa de mutação crescer (decrecer) geometricamente a razão de ω ($1/\omega$) com as gerações, caso o erro do ensemble não decresça.

O conceito que permeia o algoritmo GEFS Modificado pode ser entendido como uma taxa de mutação por demanda. Se a diversidade preditiva média não contempla os requisitos para a construção do ensemble, a taxa deve ser ajustada. Essa modificação em GEFS torna-o apropriado para construção de classificadores e eliminação dos atributos irrelevantes e redundantes presentes nos dados da LE.

Metodologia para Certificação e Validação

Qualquer algoritmo de AM pode ser utilizado com GEFS Modificado, porém resultados expressivos são obtidos com RNAs do tipo MLP treinadas por qualquer algoritmos de aprendizagem. Em especial, dois tipos algoritmos de treinamento são recomendados: o tradicional algoritmo *back-propagation* com *Early Stopping* (HAYKIN, 2001) BPES; e o algoritmo multi-objetivo MOBJ proposto por Teixeira (2000).

A estimativa de acurácia de GEFS Modificado deve ser baseada na metodologia de validação cruzada estratificada (*10-fold stratified cross-validation*) e a sua eficiência quanto ao número de atributos descartados por meio da evolução do índice de redução de características totais (IRCT³) ao longo da busca genética.

Para comprovar a superioridade ou não do algoritmo GEFS-Modificado para um determinado domínio, deve-se compará-lo em múltiplas instâncias. Recomenda-se no mínimo uma comparação pareada com um algoritmo de treinamento de um único classificador e uma auto-comparação, considerando apenas o melhor componente classificador (melhor MLP que compõe o *ensemble*). O objetivo dessas comparações é certificar a superioridade de GEFS Modificado e comprovar o postulado por Hansen e Salamon (1990).

Para comparação pareada de algoritmos, a metodologia baseada nas médias e desvios padrão em *10-fold stratified cross-validation*, descrita em Baranauskas e Monarfd (2000) deve ser adotada. Essa metodologia de comparação pode assegurar, estatisticamente, a superioridade de um algoritmo em relação a outro.

A eficiência de RCT pode ser avaliada por meio do IRCT médio a cada geração g nos r -folds de validação cruzada. Este índice é obtido por:

$$\overline{IRCT}_g^{(r)} = \frac{1}{r} \sum_{i=1}^r IRCT_i \quad (3)$$

onde $\overline{IRCT}_g^{(r)}$ é a média em r -folds dos IRCTs calculados na geração g ; e $IRCT_i$ é o índice de RCT do i -ésimo *ensemble*.

Todas as comparações e testes mencionados são suficientes para certificar a eficiência e as duas premissas teóricas de GEFS Modificado: (1) a suposta superioridade do algoritmo para construção do *ensemble* de classificadores de redes neurais em relação ao mesmo tipo de algoritmo para treinamento de um único classificador; e (2) a sua eficiência na redução de atributos dos dados.

³Percentual dos atributos eliminados

Aplicação de GEFS Modificado com a Língua Eletrônica para Classificação de Cafés

Para demonstrar o potencial de aplicação de GEFS Modificado com a LE e a referida metodologia para sua certificação e/ou validação, seis bases de dados contendo 253 medidas de cafés, igualmente distribuídas em 3 classes distintas (Ótimos, Bons e Péssimos) correspondentes à qualidade global, foram utilizadas. Cada sensor da LE gerou uma base de dados, nomeada pela letra S seguida do número de ordenamento do sensor no arranjo da LE. Além dos cinco sensores distintos, uma base foi construída por meio da junção de todos os atributos de todos os sensores ({S1, S2, S3, S4, S5}). Os parâmetros e configurações das redes MLP e do algoritmo genético foram os mesmos utilizados por Ferreira e Delbem (2006a).

Para facilitar a análise e compreensão dos resultados apresentados a seguir, algumas convenções foram estabelecidas:

- GBPES - algoritmo GEFS Modificado cujo algoritmo para treinamento das MLPs é o BPES;
- GMOBJ - algoritmo GEFS Modificado cujo algoritmo para treinamento da MLPs é o MOBJ;
- MGBES - algoritmo GBPES focado na seleção da MLP mais acurada;
- MGMEMOB - algoritmo GMOBJ focado na seleção da MLP mais acurada.

Na Tabelas 1 são apresentados os resultados obtidos com os dois diferentes algoritmos para treinamento das MLPs com GEFS Modificado para construção de *ensembles* e com uma única MLP.

Tabela 1: Estimativa de acurácia *ensembles* x RNAs bases de cafés da LE.

BASE DE DADOS	GMOBJ	GBPES	MOBJ	BPES
S1	86.10 ± 2.50	85.83 ± 1.79	62.99 ± 3.55	64.85 ± 2.20
S2	87.71 ± 2.24	89.80 ± 1.41	62.20 ± 2.90	66.50 ± 2.76
S3	86.78 ± 2.88	86.46 ± 2.75	59.76 ± 2.04	62.59 ± 2.64
S4	84.72 ± 2.43	84.58 ± 3.19	62.08 ± 3.26	62.09 ± 3.42
S5	85.63 ± 2.65	88.80 ± 2.37	62.57 ± 2.67	56.14 ± 4.10
{S1,S2,S3,S4,S5}	84.61 ± 2.22	88.31 ± 1.71	66.52 ± 1.98	65.91 ± 2.68

Para todas as bases de dados, GMOBJ e GBPES superaram, respectivamente, MOBJ e BPES, com grau de confiança de 95%.

A avaliação anterior foi estendida para MGMEMOB e MGBES. Ambos algoritmos fazem parte da mesma execução de GMOBJ e GBPES, respectivamente, mas consideram apenas a acurácia obtida pela melhor MLP do *ensemble* em cada *fold* da validação cruzada. No gráfico da Fig. 1 são apresentadas as acurácias obtidas com as bases de dados de qualidade global de cafés.

Como era esperado, GMOBJ e GBPES apresentaram maior valor da estimativa de acurácia superando, respectivamente, MGMEMOB e MGBES em todas as bases de dados. Um grau de confiabilidade de 95% dessa superação foi estatisticamente comprovado para todas as bases com o algoritmo GBPES e nas bases S3 e S4 com GMOBJ. Observa-se que GBPES é estatisticamente superior (grau confiabilidade de 95%) a GMOBJ com as bases S2, S5 e {S1,S2,S3,S4,S5}. Supõe-se que essa superioridade advinha da acuidade de ajuste dos parâmetros.

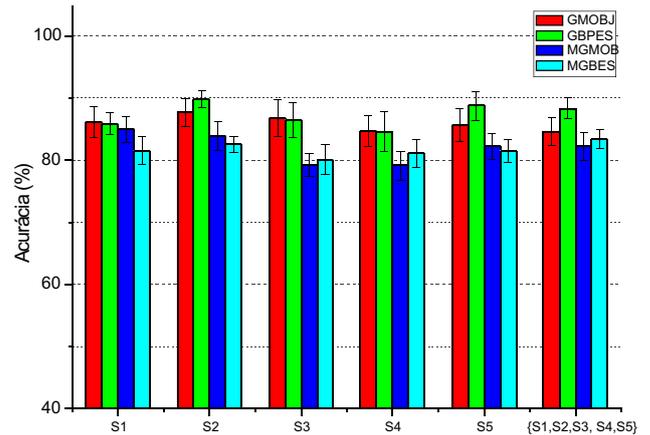


Fig. 1: Acurácia estimada dos algoritmos GMOBJ, GBPES, MGMEMOB e MGBES para os dados de cafés da LE.

Evolução do Desempenho

Embora as comparações mostradas nos gráficos anteriores sejam conclusivas a respeito da eficácia de GEFS Modificado para classificação de cafés segundo sua qualidade, uma avaliação adicional do desempenho do algoritmo durante a evolução da busca genética (evolução genética) também pode ser realizada. A evolução do desempenho é capaz de mostrar se o algoritmo refina sua acurácia (minimiza o erro) ou se sua eficácia é decorrente de uma simples amostragem randômica do espaço de busca, como ocorre em RS (HO, 1998). Logo, avaliou-se a evolução geracional da acurácia de GMOBJ e GBPES, utilizando o erro médio em validação cruzada, cálculo similar ao da Equação 3. Os resultados obtidos são apresentados nos gráficos das Fig. 2 e 3.

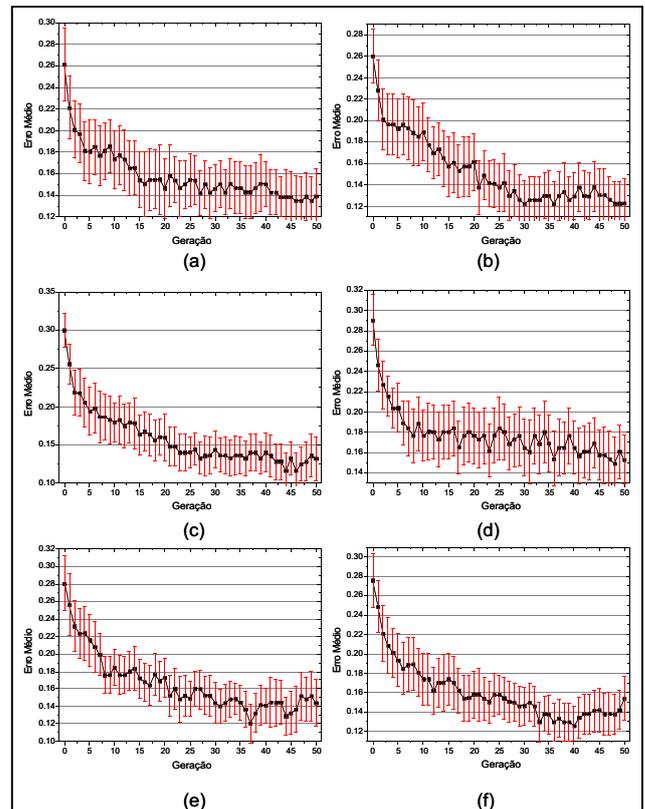


Fig. 2: Evolução do desempenho em GMOBJ: (a) S1, (b) S2, (c) S3, (d) S4, (e) S5, (f) {S1,S2,S3,S4,S5}.

A geração zero representa uma população aleatoriamente amostrada do espaço de busca, como faz RS (HO, 1998). Logo, observa-se uma tendência de decréscimo do erro, conforme o número de gerações aumenta. Essa tendência de comportamento mostra que o algoritmo GEFS Modificado refina a *ensemble* gerado durante o processo evolutivo. Por essa razão, é possível concluir que GEFS Modificado é também mais acurado do que a construção do *ensemble* por RS (HO, 1998).

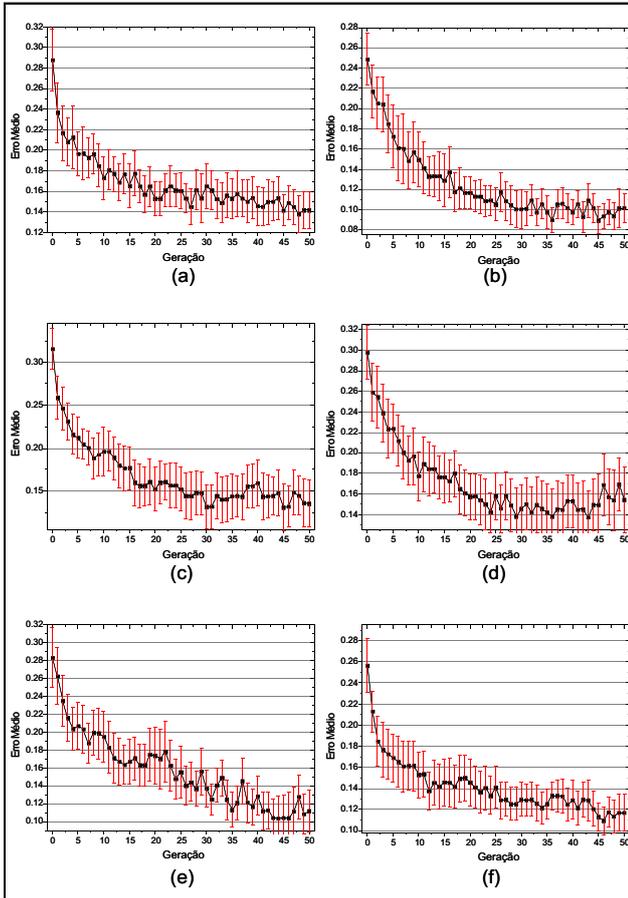


Fig. 3: Evolução do desempenho em GBPES: (a) S1, (b) S2, (c) S3, (d) S4, (e) S5 e (f) {S1,S2,S3,S4,S5}.

Eficiência em redução de atributos

A eficiência do algoritmo para seleção do subconjunto relativamente reduzido de atributos pode ser avaliada por meio da média em *r-folds* dos IRCTs, calculados a cada geração *g*. Na análise da eficiência do algoritmo para RCT, a dinâmica de evolução do algoritmo traz informações relevantes e conclusivas sobre a tendência de RCT dos algoritmos GMOBJ e GBPES. Mais do que um valor fixo obtido ao final da execução dos algoritmos, a avaliação do desempenho do algoritmo ao longo da evolução genética propicia uma interpretação sobre a tendência do processo. Por isso, a eficiência de GMOBJ e GBPES foi calculada para cada uma das gerações *g* por meio do $\overline{IRCT}_g^{(r)}$ nos 10 *folds* de validação cruzada. Na Fig. 4 são mostrados os gráficos com a evolução para todas as gerações e bases consideradas na execução dos algoritmos GMOBJ e GBPES.

Observa-se, por meio dos gráficos da Fig. 4, uma clara tendência de crescimento do IRCT médio ao longo da busca genética. Esse comportamento foi similar para

todas as bases consideradas. Conforme o algoritmo evolui, geração a geração, atributos irrelevantes no processo de construção do *ensemble* são descartados. Isso comprova a eficiência de GEFS Modificado como algoritmo para construção de um *ensemble* e seleção de atributos dos dados provenientes de uma LE.

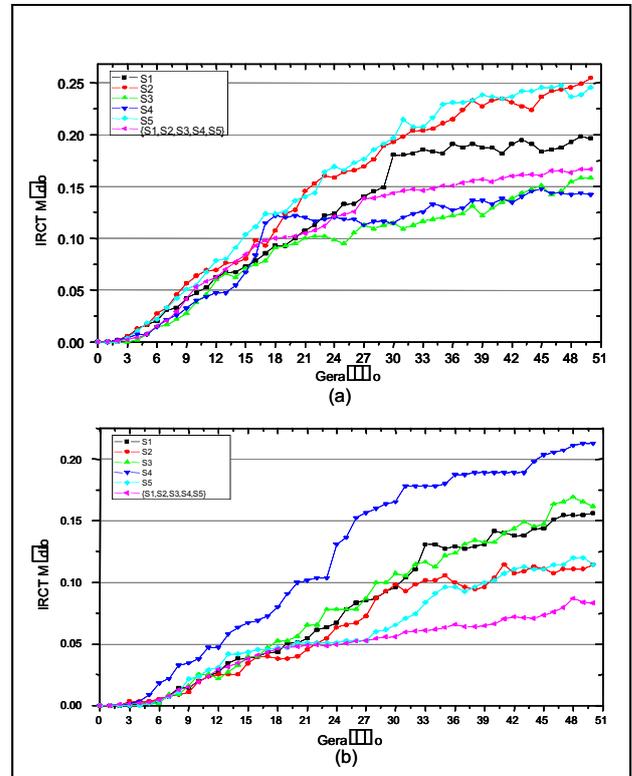


Fig. 4: Evolução do IRCT médio: (a) algoritmo GMOBJ e (b) algoritmo GBPES.

Conclusões

Os resultados apresentados na seção anterior são motivadores e comprovam a eficácia do algoritmo GEFS Modificado para construção de um *ensemble* de RNAs no domínio da LE. As comparações pareadas entre o algoritmo para treinamento de uma única RNAs e o mesmo algoritmo visando o melhor indivíduo da população deflagram e atestam a sua superioridade. No que tange à questão da evolução da acurácia do *ensemble*, constatou-se uma tendência de decréscimo do erro de classificação à medida que o número de gerações aumenta. O algoritmo GEFS Modificado preserva, portanto, as características primordiais de construção de *ensembles* do algoritmo GEFS original e efetua ainda seleção de atributos, requisito fundamental no domínio da LE.

Ambas abordagens (MOBJ e BPES) para treinamento das redes MLP aplicadas são adequadas para uso com GEFS Modificado nas bases geradas pela LE. Com o algoritmo BPES os resultados obtidos foram ligeiramente melhores e com maiores significâncias para algumas bases. Provavelmente essa superioridade advinha de maior acuidade de ajustes realizada durante as definições paramétricas. No entanto, as duas abordagens são recomendadas para uso com GEFS Modificado.

A seleção de atributos realizada por GEFS Modificado é

um norteador para escolha das frequências relevantes de operação dos sensores da LE. Um *ensemble* de RNAs construído com GEFS Modificado para classificação de cafés, por exemplo, além de elevar a acurácia do classificador, possibilita a seleção dos atributos (frequências para os sensores) mais relevantes, evitando assim a construção de equipamentos de maior complexidade eletrônica e minimizando o custo, o desperdício de armazenamento e o tempo demandados na coleta do dados. GEFS Modificado também se mostrou eficaz para utilização com um único sensor da LE para classificação de cafés. Uma acurácia aproximada de 90 % foi obtida com apenas um sensor e com um subconjunto de frequências menor que 80% do total. Esse é um dos aspectos de maior relevância, pois revela a possibilidade de construção de uma LE para classificação de cafés, segundo sua qualidade global, ainda mais simples, com apenas um sensor e com um conjunto reduzido de frequências. Por essa razão GEFS Modificado, bem como a metodologia para sua certificação e validação descrita nesta circular, estão sendo utilizadas nos dados obtidos com a LE.

Referências

- BARANAUSKAS, J. A.; MONARFD, M. C. **Reviewing some machine learning concepts and methods**. São Carlos: USP-ICMC, 2000. 52 p. (USP-ICMC. Relatórios técnicos, 102).
- BERNARDINI, F. C. **Combinação de classificadores simbólicos para melhorar o poder preditivo e descritivo de ensembles**. 2002. 85 f. Dissertação (Mestrado em Ciência da Computação) Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.
- BREIMAN, L. Bagging predictors. **Machine Learning**, Boston, v. 24, n. 2, p. 123-140, 1996.
- DIETTERICH, T. G. *Ensemble methods in machine learning*. In: INTERNATIONAL WORKSHOP ON MULTIPLE CLASSIFIER SYSTEMS, 1., 2000. **Anais...** New York: Springer Verlag, 2000. p. 1-15. (Lecture notes in computer science). Disponível em: <<http://www.cs.orst.edu/~tgd/>>.
- FERREIRA, E. J.; DELBEM, A. C. B. Abordagem genética para seleção de um conjunto reduzido de características para construção de *ensembles* de redes neurais: aplicação à língua eletrônica. In: SIMPÓSIO DE TESES E DISSERTAÇÕES, 10., 2006, São Carlos, SP. **Anais...** [S. l.: s. n.], 2006a. 1 CD-ROM. não paginado.
- FERREIRA, E. J.; DELBEM, A. C. B. A new approach based on genetic *ensemble* feature selection to construct neural network *ensembles*: an application to gustative sensors. In: INTERNATIONAL JOINT CONFERENCE WTDIA 06, 2006, Ribeirão Preto, SP. **Anais...** [S. l.: s. n.], 2006b. 1 CD-ROM. não paginado.
- FREUND, Y.; SCHAPIRE, R. E. Experiments with a new boosting algorithm. In: SAITTA, L. (Ed.). **ICML'96: proceedings of the 13th International Conference on Machine Learning**, 1996, San Francisco. San Francisco: Morgan Kaufmann, 1996. p. 146-148.
- HANSEN L.; SALAMON, P. Neural network *ensembles*. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 12, p. 993-1001, 1990.
- HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Porto Alegre: Bookman, 2001. 900 p.
- HO, T. K. The random subspace method for construction decision forests. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 20, n. 8, p. 832-844, 1998.
- KOLLER, D.; SAHAMI, M. Toward Optimal Feature Selection. In: PROCEEDING OF 13th INTERNATIONAL CONFERENCE IN MACHINE LEARNING, 1996, Bari, Itália. **Proceedings...** [S. l.: s. n., 1996]. p. 284-292. Disponível em: <<http://citeseer.ist.psu.edu/koller96toward.html>>.
- OPITZ, D. W. Feature selection for *ensembles*. In: INTERNATIONAL JOINT CONFERENCE ON ARTIFICIAL INTELLIGENCE, 16., 1999, STOCKHOLM. **Proceedings...** Somerset: International Joint Conferences on Artificial Intelligence, Inc., 1999. p. 379-384. Disponível em: <<http://citeseer.nj.nec.com>>. Acesso em: 15 jan. 2004.
- OPITZ, D. W.; SHAVLIK, J. W. A genetic algorithm approach for creating neural-network *ensembles*. In: SHARKEY, A. (Ed.). **Appears in combining artificial neural nets**. London: Springer-Verlag, 1999. p. 79-97.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representation by back-propagating errors. **Nature**, London, v. 323, p. 533-536, 1986.
- RUSSELL, S.; NORVIG, P. **Inteligência artificial**. 2. ed. Rio de Janeiro: Elsevier, 2004. 1021 p. ISBN: 85-352-1177-2. Tradução de Vandenberg D. de Souza. Consultoria editorial e revisão técnica de Raul Sidnei Wazlawick.
- TEIXEIRA, R. de A. et. al. Improving generalization of MLPs with multi-objective optimization. **Neurocomputing**, [S. l.], v. 35, p. 189-194, 2000.
- TSYMBAL, A.; PECHENIZKIY, M.; CUNNINGHAM, P. **Diversity in ensemble feature selection**. Technical report, Trinity College Dublin, 2003a. não paginado. Disponível em: <<https://www.cs.tcd.ie/publications/tech-reports/reports.03/TCD-CS-2003-44.pdf>>. Acesso em: 13 fev. 2004.
- TSYMBAL, A. et. al. Search strategies for *ensemble* feature selection in medical diagnostics. In: 16th IEEE SYMPOSIUM ON COMPUTER-BASED MEDICAL SYSTEMS, 16., 2003, New York. **Proceedings...** New York, IEEE CS Press, 2003b. p. 124-129.

Circular Técnica, 34

Ministério da Agricultura,
Pecuária e Abastecimento

Exemplares desta edição podem ser adquiridos na:
Embrapa Instrumentação Agropecuária
Rua XV de Novembro, 1542 - Caixa Postal 741
CEP 13560-970 - São Carlos-SP
Fone: 16 3374 2477
Fax: 16 3372 5958
E-mail: sac@cnpdia.embrapa.br
www.cnpdia.embrapa.br

1a. edição

1a. impressão 2006: tiragem 300

Comitê de Publicações

Presidente: Dr. Carlos Manoel Pedro Vaz
Membros: Dra. Débora Marcondes B. P. Milori,
Dr. João de Mendonça Naime,
Dr. Washington Luiz de Barros Melo
Valéria de Fátima Cardoso

Membro Suplente: Dr. Paulo S. P. Herrmann Junior

Expediente

Revisor editorial: Dr. Victor Bertucci Neto
Normalização bibliográfica: Valéria de Fátima Cardoso
Tratamento das ilustrações: Valentim Monzane
Foto da capa: Ednaldo J. Ferreira
Editoração eletrônica: Valentim Monzane