

Estudo Comparativo entre Critérios de Busca para Seleção Estática de Conjuntos de Classificadores

Eulanda M. Dos Santos

¹Universidade Federal do Amazonas

{emsantos}@dcc.ufam.edu.br

Abstract. Classifier ensemble selection is an optimization process frequently conducted using non-exhaustive search algorithm. However, the problem of choosing the most appropriate search criterion is an open problem in the literature. Performance is the most obvious search criterion since it allows us to accomplish the main objective in classification problems, which is to find high-performance predictors. On the other hand, it is widely accepted that diversity is an important criterion in classifier ensembles. Even though, the relationship between diversity and performance is unclear. Finally, since classifier members are trained to solve the same classification problem, there are redundant classifiers among them. Therefore, it is possible to increase recognition rates while reducing the number of classifiers to meet both the performance and complexity requirements. We present in this paper a comparative study in order to point out the best objective function for finding high-performance classifier ensembles. The search criteria investigated are: ensemble error rate, ensemble size and diversity measures. First, the search criteria are directly compared in a single-objective optimization approach performed by single-objective genetic algorithm (GA). Then, search criteria are combined to make up pairs of objective functions to guide a multi-objective optimization process performed by multi-objective GA.

Resumo. A seleção de subconjuntos de classificadores é um processo de otimização que pode ser realizado por algoritmos de busca, normalmente não exaustivos. Porém, a escolha do critério mais apropriado para guiar a busca é um problema ainda não solucionado na literatura. Por um lado, o principal objetivo de métodos de classificação é uma elevada taxa de acerto. Por outro lado, é intuitivamente aceito que os membros do conjunto devem apresentar diversidade de opinião, especialmente diversidade de erros. Entretanto, a relação entre diversidade e taxa de classificação não é clara na literatura. Por fim, como os membros do conjunto são treinados para resolver o mesmo problema de classificação, é elevada a probabilidade de redundância entre os mesmos. Logo, o tamanho do conjunto pode ser reduzido através da eliminação de classificadores membros redundantes, aumentando ao mesmo tempo a taxa de acerto. Neste artigo é feito um estudo comparativo entre taxa de acerto, medidas de diversidade e tamanho dos subconjuntos como critérios de busca para a seleção de subconjuntos de classificadores. Inicialmente, os critérios de busca são comparados diretamente utilizando-se algoritmo genético mono-objetivo. Em um segundo grupo de experimentos, os critérios de busca são combinados em pares de funções objetivo para guiar um algoritmo genético multi-objetivo. O objetivo do artigo é identificar a melhor função objetivo ou par de funções objetivo para o problema investigado.

1. Introdução

A técnica de conjuntos de classificadores tornou-se uma estratégia dominante em muitas áreas de aplicação como aprendizagem de máquina, reconhecimento de padrões e *data mining*. A razão desse sucesso são os estudos experimentais [Ho 1998] e teóricos [Dietterich 2000] que mostram que conjuntos de classificadores podem ter taxas de acerto superiores às taxas obtidas por classificadores individuais. Uma vez criado um conjunto de classificadores, com técnicas como *bagging* [Breiman 1996] e subespaços aleatórios [Ho 1998], a estratégia mais comum é a combinação da saída de todos os classificadores. Essa estratégia assume que todos os membros são igualmente importantes e que apresentam decisões independentes. Por outro lado, a seleção de classificadores parte do princípio de que os membros do conjunto inicial são redundantes [Zhou et al. 2002], ou seja, esta estratégia busca encontrar o mais relevante subconjunto de classificadores, para então combinar unicamente a saída dos membros do subconjunto selecionado.

A seleção de subconjuntos de classificadores, conhecida na literatura como SOCS (do Inglês *static overproduce-and-choose strategy*), é dividida em duas fases: (1) super produção e (2) seleção. Na primeira fase, muitos classificadores são treinados para compor o conjunto inicial. O conjunto criado pode ser homogêneo, isto é, os membros são baseados no mesmo algoritmo de aprendizagem, ou heterogêneos, em que os classificadores são baseados em diferentes algoritmos como redes neurais artificiais, *Support Vector Machines* (SVM), *k Nearest Neighbors* (kNN), dentre outros. Na segunda fase de SOCS ocorre o teste de diferentes combinações dos membros do conjunto inicial para que seja identificado o melhor subconjunto de classificadores. Portanto, a segunda fase é a etapa mais importante da técnica SOCS [Ruta and Gabrys 2005].

Existem dois parâmetros fundamentais na etapa de seleção de SOCS: algoritmo de busca e função objetivo. Vários algoritmos de busca têm sido aplicados, tais como ordenamento dos n melhores classificadores, Algoritmos Genéticos (GA) [Ruta and Gabrys 2005] e *hill-climbing* [Zenobi and Cunningham 2001]. A escolha da função objetivo, porém, ainda é uma questão não resolvida na literatura. Taxa de acerto elevada é o critério mais importante em problemas de classificação em geral, porém, é amplamente aceito na literatura que diversidade entre os membros do conjunto é fundamental para que um conjunto supere classificadores individuais. Embora esses dois critérios sejam fundamentais ao sucesso de conjuntos de classificadores, a relação entre taxa de acerto e diversidade não é clara [Kuncheva and Whitaker 2002]. Além disso, considerando que SOCS baseia-se no fato de que os componentes do conjunto são redundantes, é possível também usar como função objetivo a eliminação dos membros redundantes, elevando dessa forma a taxa de classificação e reduzindo ao mesmo tempo a complexidade do conjunto.

Esses três critérios de busca já foram investigados na literatura, individualmente [Aksela and Laaksonen 2006], [Ruta and Gabrys 2005] ou combinados [Zenobi and Cunningham 2001] e [Tremblay et al. 2004]. Porém, além de não haver um consenso nos resultados, poucas medidas de diversidade foram investigadas, sendo que há muitas medidas que podem ser aplicadas como função objetivo. É também importante destacar que não há estudos combinando diversidade e tamanho dos subconjuntos como função objetivo.

Diante desse contexto, o objetivo deste artigo é apresentar um abrangente es-

tudo experimental para identificar o critério de busca que possibilite encontrar o subconjunto de classificadores com maior desempenho, isto é, com maior taxa de classificação. As funções objetivo são comparadas aplicando-se algoritmos genéticos mono e multi-objetivo. São aplicadas 14 medidas: 12 medidas de diversidade, além de taxa de acerto e tamanho dos subconjuntos. As medidas de diversidade são comparadas diretamente com taxa de acerto, como critérios para guiar GA mono-objetivo. Posteriormente, os critérios de busca são combinados em pares de funções objetivo para guiar GA multi-objetivo. Nesse segundo grupo de experimentos, os pares de funções objetivo são compostas por diversidade combinada com taxa de acerto, diversidade combinada com tamanho dos subconjuntos e taxa de acerto combinada com tamanho dos subconjuntos.

O restante deste artigo está organizado da seguinte forma. Na seção 2 é descrito o processo de seleção de subconjuntos de classificadores e são destacados alguns trabalhos relacionados. A seção 3 descreve alguns dos critérios de busca investigados nos experimentos. Por fim, os resultados obtidos nos experimentos são apresentados e discutidos na seção 4. Conclusões e sugestões para trabalhos futuros são discutidos na seção 5.

2. Seleção de Subconjuntos de Classificadores

A figura 1 ilustra o processo de seleção de subconjuntos de classificadores que ocorre da seguinte forma: um conjunto de classificadores $\mathcal{C} = \{c_1, c_2, \dots, c_n\}$ é criado através de um método de geração de conjuntos de classificadores, os quais são treinados com amostras de uma base de treinamento \mathcal{T} . Posteriormente, um processo de otimização ocorre para testar diferentes subconjuntos candidatos utilizando-se uma base de otimização \mathcal{O} . Por fim, a melhor solução C_j^* encontrada durante a otimização é então usada para classificar amostras desconhecidas, que formam a base de teste ou generalização \mathcal{G} . Para que o processo de otimização não seja afetado pelo problema de aprendizagem viciada, é necessário que uma quarta base de dados seja usada. Essa base de dados independente é conhecida na literatura como base de validação \mathcal{V} .

Aprendizagem viciada é um importante problema que afeta muitas aplicações de classificação. Trata-se do fenômeno detectado quando o algoritmo de aprendizagem adapta-se demasiadamente ao conjunto de treinamento a ponto de memorizar ruídos e peculiaridades da referida base. Como consequência, ocorre uma generalização inferior ao desejado. Isso quer dizer que o sistema de classificação apresentará baixa taxa de reconhecimento quando amostras não conhecidas forem testadas.

Algoritmos de busca são usados com frequência quando o conjunto \mathcal{C} possui muitos membros devido à complexidade exponencial que uma busca exaustiva demandaria, uma vez que, sendo $\mathcal{P}(\mathcal{C})$ o conjunto de todos os subconjuntos possíveis dos membros de \mathcal{C} , o tamanho de $\mathcal{P}(\mathcal{C})$ é 2^n , sendo n o número de classificadores em \mathcal{C} . Além disso, a fase de seleção de OCS pode ser facilmente formulada como um problema de otimização em que o algoritmo de busca minimiza/maximiza uma função objetivo ou um grupo de funções objetivo.

O problema da escolha do critério de busca mais apropriado é o alvo de alguns trabalhos encontrados na literatura nos últimos anos. Porém, os resultados desses estudos não convergem para um consenso, embora todos concordem que diversidade de opinião entre os membros do conjunto é de fundamental importância. Conjuntos de classificadores superam a taxa de reconhecimento de classificadores individuais apenas

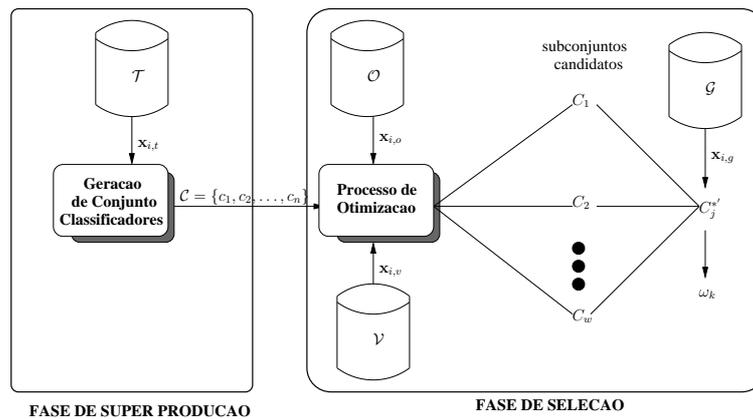


Figura 1. Fases de super produção e seleção em SOCS. A fase de seleção é formulada como um processo de otimização, que gera diferentes subconjuntos candidatos. Uma base de validação é usada no processo de otimização para evitar o problema de aprendizagem viciada. Por fim, a melhor solução escolhida é selecionada para classificar as amostras de teste.

quando ocorre diversidade entre eles [Kuncheva and Whitaker 2002]. Aksela e Laaksonen [Aksela and Laaksonen 2006] aplicaram com sucesso um método de seleção de subconjuntos de classificadores com critério de busca baseado em diversidade de erros. Entretanto, a relação entre diversidade e taxa de reconhecimento não é clara. Kuncheva e Whitaker [Kuncheva and Whitaker 2002] mostraram que diversidade e taxa de acerto não possuem uma forte relação, e concluíram que o cálculo da taxa de acerto não pode ser substituído por diversidade. Esses resultados foram confirmados por Ruta e Gabrys [Ruta and Gabrys 2005] no contexto de seleção de subconjuntos de classificadores. Eles usaram medidas de diversidade para guiar a seleção de subconjuntos de classificadores no intuito de reduzir a taxa de erro de classificação. Eles concluíram que diversidade não é melhor do que taxa de acerto.

A combinação de taxa de acerto com diversidade pode ser uma forma de superar esse aparente dilema, pois permite o uso simultâneo dessas medidas como critério de busca. Não é surpresa o fato de que essa alternativa já foi verificada na literatura. Opitz e Shavlik [Opitz and Shavlik 1996] aplicaram GA usando uma única função objetivo que combinava taxa de acerto e diversidade (conforme definido em [Opitz and Shavlik 1996]) na seleção de subconjuntos de redes neurais artificiais. Eles concluíram que essa estratégia superou a simples combinação dos classificadores iniciais. Zenobi e Cunningham [Zenobi and Cunningham 2001] criaram conjuntos de kNN aplicando seleção de características usando ambiguidade (conforme definida em [Zenobi and Cunningham 2001]) e taxa de acerto como funções objetivo. O método de busca usado foi *hill-climbing*. Os autores mostraram que essa combinação de diversidade e taxa de acerto superou o uso de taxa de acerto como única função objetivo. Tremblay et al. [Tremblay et al. 2004] também aplicaram a combinação de taxa de acerto com ambiguidade como um par de funções objetivo, enquanto GA multi-objetivo (MOGA) foi usado como algoritmo de busca. Eles concluíram, porém, que MOGA guiado por esse par de funções objetivo não encontrou soluções melhores do que as soluções encontradas por GA usando unicamente taxa de acerto. É importante destacar que ambiguidade foi a única medida de diversidade investigada nesses três trabalhos. Porém, existem muitas outras medidas de diversidade

propostas na literatura [Kuncheva and Whitaker 2002].

Portanto, taxa de acerto (ou erro de classificação), tamanho dos classificadores e medidas de diversidade são os critérios de busca mais usados na literatura. O primeiro é o critério mais óbvio pois tem como foco a redução direta do erro de classificação. Em termos de tamanho dos subconjuntos, a minimização do número de classificadores membros permite reduzir a complexidade das soluções e ao mesmo tempo aumentar a taxa de acerto. Por fim, o papel fundamental da diversidade é claramente definido na literatura. Esses três critérios são usados neste trabalho e descritos na próxima seção.

3. Critérios de Busca Investigados

Conforme mencionado anteriormente, são investigados neste artigo três critérios de busca: (1) tamanho dos subconjuntos; (2) taxa de erro; e (3) diversidade. A função objetivo relacionada com o tamanho dos subconjuntos é definida como o número de classificadores membros do subconjunto. O segundo critério, taxa de erro (ϵ), relaciona-se ao erro obtido pelas diferentes combinações de classificadores membros. Logo, uma função de combinação deve ser usada. Dentre as diferentes possibilidades de funções, o simples voto majoritário é usado como função de combinação neste trabalho. Essa função é definida da seguinte forma: dado um conjunto com n classificadores, y_i como a saída do i -ésimo classificador e o conjunto $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ de possíveis rótulos de classes do problema, voto majoritário (mv) é calculado como:

$$mv = \max_{k=1}^c \sum_{i=1}^n y_{i,k} \quad (1)$$

Diversidade constitui o terceiro critério de busca investigado neste trabalho. A definição de diversidade está relacionada com o fato de que os membros do conjunto devem cometer erros diferentes uns dos outros. Diferentes medidas de diversidade foram propostas na literatura. Neste trabalho serão investigadas 12 medidas. Por questão de limitação de espaço nem todas serão descritas aqui. Apenas as medidas mais diferentes entre si são apresentadas nesta seção. Porém, a tabela 1 lista todas as medidas usadas nos experimentos e as referências respectivas.

Para que as definições das medidas de diversidade sejam apresentadas, é usada a seguinte notação. Dado C_j um subconjunto de classificador candidato, X a base de dados e n e l seus respectivos tamanhos. N^{ab} representa o número de amostras classificadas em X , em que a, b podem assumir o valor 1 quando o classificador estiver correto e 0 caso contrário. $r(x)$ representa o número de classificadores que classificam corretamente a amostra x . É importante destacar que medidas de dissimilaridade devem ser maximizadas, enquanto medidas de similaridade devem ser minimizadas quando usadas como função objetivo. Além disso, as medidas *pairwise* são calculadas para cada par de classificadores c_i e c_k , enquanto as medidas não *pairwise* são calculadas no subconjunto C_j completo.

1. **Ambiguidade:** Dado o subconjunto candidato $C_j = \{c_1, c_2, \dots, c_l\}$ e a saída produzida por ele ω_k , a ambiguidade é calculada conforme equação 2:

$$a_i(x) = \begin{cases} 0 & \text{if } y_i = \omega_k \\ 1 & \text{caso contrário} \end{cases} \quad (2)$$

o valor $a_i(x)$ é a ambiguidade do i -ésimo classificador para a amostra x . Considerando X a base de dados, $|X|$ sua quantidade de amostras, e $|C_j|$ o tamanho do subconjunto, a ambiguidade para todo o subconjunto C_j é:

$$\gamma = \frac{1}{|X| \cdot |C_j|} \sum_{i \in C_j} \sum_{x \in X} a_i(x) \quad (3)$$

2. **Medida de Dificuldade:** Dado F obtido a partir de $\{\frac{0}{l}, \frac{1}{l}, \dots, 1\}$, que representa o número de classificadores em C_j que classificam corretamente x , e $Var(F)$ a variância em F , esta medida é calculada da seguinte forma:

$$\theta = Var(F) \quad (4)$$

3. **Falha Coincidente:** Esta medida é baseada na mesma distribuição F usada para calcular θ . Aqui porém, $Y = \frac{i}{l}$ representa a proporção de classificadores que não classificam corretamente x . Considerando $p(q)$ como a probabilidade que q classificadores erram ao classificar x , *Coincident Failure* é definida como:

$$\sigma = \begin{cases} 0, & p_0 = 1.0 \\ \frac{1}{(1-p_0)} \sum_{q=1}^l \frac{q(q-1)}{l(l-1)} p_q & p_0 < 0 \end{cases} \quad (5)$$

4. **Falha Dupla:** Assumindo N^{ab} como a quantidade de amostras classificadas em X , sendo que a, b determinam se o classificador está correto (1) ou não (0). Esta medida *pairwise* é definida para o par de classificadores c_i e c_k da seguinte forma:

$$\delta_{i,k} = \frac{N^{00}}{N^{11} + N^{10} + N^{01} + N^{00}} \quad (6)$$

Embora apenas 04 medidas de diversidade estejam descritas nesta seção, 12 medidas foram usadas nos experimentos. A tabela 1 mostra essas medidas, além de destacar que taxa de erro e tamanho dos subconjuntos também foram investigados como critério de busca. A descrição dos experimentos e os resultados obtidos são apresentados na seção 4.

4. Experimentos

Os experimentos foram realizados com um conjunto de classificadores constituído por membros do tipo kNN, sendo que o valor de k foi definido como $k = 1$, sem que tenha sido feito ajuste do parâmetro para evitar experimentos adicionais. O método de subespaços aleatórios [Ho 1998] foi usado para gerar um conjunto de 100 kNNs. Este método funciona da seguinte forma: uma parcela de características iniciais é escolhida aleatoriamente e usada para representar os dados de treinamento de um classificador. Essa parcela de características é chamada de subespaço. O processo de escolha de subespaços é repetido n vezes, sendo que n representa o número de classificadores membros do conjunto.

Foi utilizada a base de dados NIST SD19 *digits Special Database*, que é uma base de dados usada com muita frequência em estudos sobre algoritmos de reconhecimento de dígitos manuscritos. Essa base possui muitas amostras para possibilitar a divisão nas quatro partições necessárias, isto é: treino, otimização, validação e teste, e um grande

Tabela 1. Lista dos critérios de busca usados nos experimentos. As referências indicam os autores de cada medida. A coluna tipo indica se o critério de busca deve ser minimizado (similaridade) ou maximizado (dissimilaridade)

Nome	Símbolo	Tipo
Taxa de erro	ϵ	Similaridade
Tamanho dos subconjuntos	ζ	Similaridade
Ambiguidade	γ [Zenobi and Cunningham 2001]	Dissimilaridade
Falha Coincidente	σ [Kuncheva and Whitaker 2002]	Dissimilaridade
Coefficiente de Correlação	ρ [Kuncheva and Whitaker 2002]	Similaridade
Medida de Dificuldade	θ [Kuncheva and Whitaker 2002]	Similaridade
Desacordo	η [Kuncheva and Whitaker 2002]	Dissimilaridade
Falha Dupla	δ [Kuncheva and Whitaker 2002]	Similaridade
Entropia	ξ [Kuncheva and Whitaker 2002]	Dissimilaridade
Falha Majoritária	λ [Ruta and Gabrys 2005]	Dissimilaridade
Diversidade Generalizada	τ [Kuncheva and Whitaker 2002]	Dissimilaridade
Acordo entre Médias	κ [Kuncheva and Whitaker 2002]	Similaridade
Kohavi-Wolpert	ψ [Kuncheva and Whitaker 2002]	Dissimilaridade
Q-estatístico	Φ [Kuncheva and Whitaker 2002]	Similaridade

Tabela 2. Especificações da base de dados usadas nos experimentos.

Base de dados	Número de características	Base de Treino (\mathcal{T})	Base de Otimização (\mathcal{O})	Base de Validação (\mathcal{V})	Base de Teste (\mathcal{G})
NIST-digits	132	5,000	10,000	10,000	test1 60,089 test2 58,646

espaço de características necessário ao método de subespaços aleatórios. A base foi dividida nas quatro partições independentes usando o método clássico de validação cruzada conhecido na literatura com o *holdout validation*. Originalmente há duas partições de teste da base NIST-digits, chamadas aqui de *data-test1* (60,089 amostras) e *data-test2* (58,646 amostras). A base *data-test2* é composta por amostras mais difíceis de classificar [Gother 1995]. Na tabela 2 são listadas informações importantes sobre as partições. Por fim, foi aplicada a representação proposta por Oliveira et al. [Oliveira et al. 2002]. Essa representação é composta por 132 características, sendo 78 de concavidade, 48 de contorno e 6 de superfície. Para o método de subespaços aleatórios, foram utilizados subespaços com 32 características.

Dois algoritmos de busca foram aplicados: GA e MOGA. O algoritmo do tipo MOGA usado neste artigo foi NSGA-II (*elitist non-dominated sorting genetic algorithm*) [Deb 2001]. Essa escolha deve-se a duas características importantes de NSGA-II: estratégia de elitismo completo e um mecanismo de preservação de diversidade de indivíduos baseada em cálculo de distância que não necessita de ajuste de parâmetros. Para seguir o mesmo protocolo experimental com os dois algoritmos de busca investigados, os mesmos valores de parâmetros foram usados. O processo de otimização foi realizado por GA e MOGA usando vetores binários. Uma vez que o conjunto de classificadores inicial contém 100 membros, cada indivíduo foi representado por um vetor binário com

Tabela 3. Parâmetros dos Algoritmos Genéticos

Tamanho da população	128
Número de gerações	1000
Probabilidade de cruzamento	0.8
Probabilidade de mutação	0.01
Cruzamento <i>One-point</i> e Mutação <i>bit-flip</i>	

100 posições. Cada bit determina a presença (1) ou a ausência (0) do classificador. Os parâmetros dos algoritmos de busca estão resumidos na tabela 3.

4.1. Resultados

Todos os experimentos foram replicados 30 vezes e os resultados foram testados em comparações múltiplas usando o teste estatístico não paramétrico Kruskal-Wallis. O nível de confiança foi 95% ($\alpha = 0.05$), e a correção Dunn-Sidak foi usada para valores críticos. Portanto, os resultados aqui apresentados são valores médios dos resultados obtidos nas 30 replicações. Os resultados estão divididos em dois grupos, análise de desempenho e análise do tamanho dos subconjuntos.

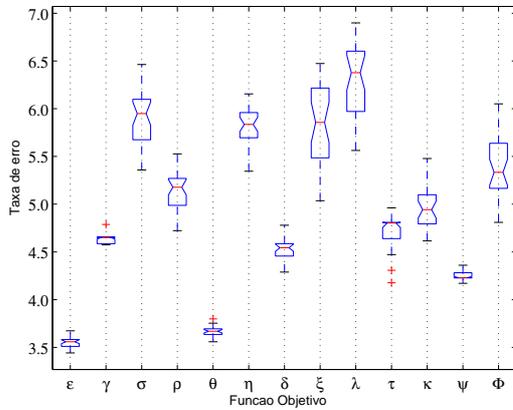
4.2. Análise de Desempenho

Primeiramente, 13 critérios de busca foram comparados diretamente, sendo a taxa de erro ϵ (1–taxa de acerto) e as 12 medidas de diversidade. Cada critério de busca foi usado como uma função objetivo para guiar GA mono-objetivo. O tamanho dos subconjuntos não foi usado como função objetivo porque o resultado produziria subconjuntos de classificadores com poucos membros. Essa comparação direta permite avaliarmos se diversidade pode ser melhor que taxa de erro na seleção de subconjuntos de classificadores.

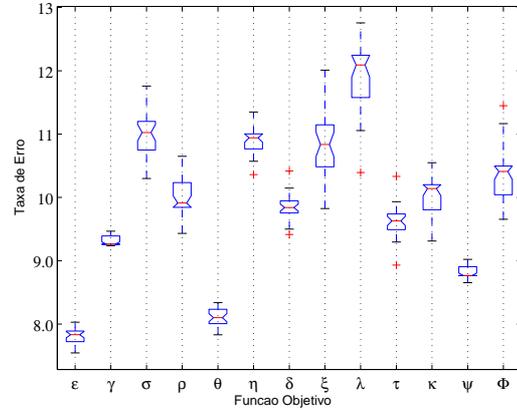
A figura 2 mostra os resultados da comparação considerando as 30 replicações nas bases *data-test1* (figura 2(a)) e *data-test2* (figura 2(b)). As figuras mostram que as medidas de diversidade não superam a taxa de classificação obtida pelas soluções encontradas utilizando-se a taxa de erro como função objetivo. Esses resultados já eram esperados, pois confirmam os resultados obtidos em trabalhos relacionados [Ruta and Gabrys 2005] e [Kuncheva and Whitaker 2002]. Uma vez confirmada que diversidade não é melhor do que ϵ , precisamos testar se a combinação desses dois critérios de busca em pares de funções objetivo pode melhorar os resultados obtidos por ϵ .

Cada medida de diversidade foi combinada com ϵ para compor pares de funções objetivo a fim de guiar o processo de otimização. A figura 3 mostra os resultados das 30 replicações nas bases *data-test1* (figura 3(a)) e *data-test2* (figura 3(b)). É importante destacar que o primeiro valor corresponde aos resultados obtidos por GA como algoritmo de busca e ϵ como função objetivo. Isso significa que o melhor resultado obtido na seleção mono-objetiva está sendo comparado aos resultados obtidos na seleção multi-objetiva. O segundo valor corresponde aos resultados obtidos por NSGA-II guiado por tamanho dos subconjuntos ζ e ϵ (resultados discutidos na próxima seção).

Esses resultados indicam que, ao compor pares de funções objetivo combinando diversidade e ϵ em um processo de otimização multi-objetivo, é possível encontrar subconjuntos de classificadores com desempenho, taxa de acerto, superiores às soluções en-

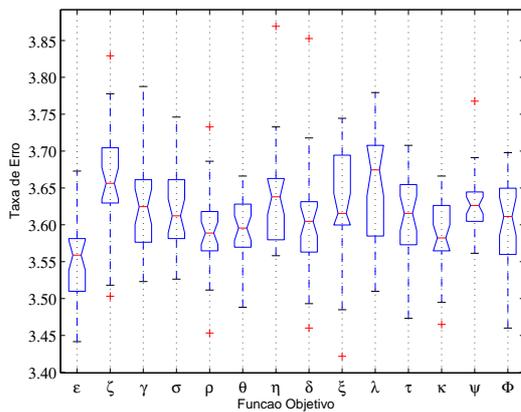


(a) data-test1

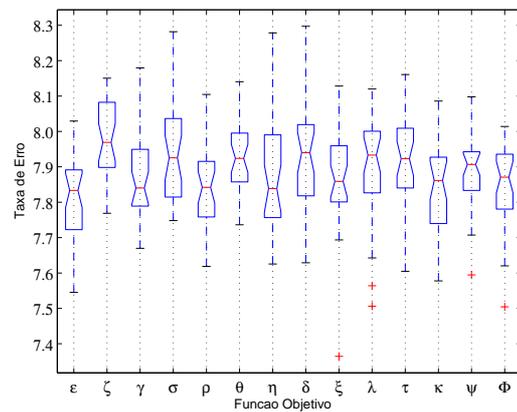


(b) data-test2

Figura 2. Resultados de 30 replicações usando GA e 13 diferentes funções objetivo. O desempenho foi calculado nas bases *data-test1* (figura 2(a)) e *data-test2* (figura 2(b)).



(a) data-test1



(b) data-test2

Figura 3. Resultados de 30 replicações usando NSGA-II e 13 diferentes pares de funções objetivo. O desempenho foi calculado nas bases *data-test1* (figura 3(a)) e *data-test2* (figura 3(b)). O primeiro valor corresponde a GA e taxa de erro ϵ como função objetivo, enquanto que o segundo valor corresponde a NSGA-II guiado por ϵ com tamanho dos subconjuntos ζ .

contradas utilizando-se unicamente diversidade. A diferença entre as soluções encontradas pelos pares de funções objetivo e as soluções encontradas usando unicamente ϵ (otimização mono-objetiva) foi bastante reduzida. Os subconjuntos de classificadores encontrados pelas três melhores medidas de diversidade foram, em média, 0.05% piores do que os encontrados por ϵ na base *data-test1* e 0.13% na base *data-test2*. Porém, os pares de funções objetivo não superam os resultados obtidos com taxa de erro como única função objetivo.

Foi mencionado na introdução que, além de taxa de erro e medidas de diversidade, o tamanho dos subconjuntos de classificadores também é importante. Esse terceiro critério de busca é investigado na próxima seção.

4.3. Análise de Tamanho dos Subconjuntos

O critério de busca relacionado ao tamanho dos subconjuntos de classificadores ζ deve ser combinado com outros critérios de busca na esperança de aumentar o desempenho e reduzir o número de classificadores. Diante desse contexto, foram criados pares de funções objetivo através da combinação de ζ com ϵ e com cada uma das 12 medidas de diversidade. Esses pares de funções objetivo foram usadas para guiar o processo de busca realizado por NSGA-II. A figura 4 mostra o tamanho dos subconjuntos encontrados por cada par de função objetivo. A taxa de classificação obtida pelas soluções são mostradas na figura 5 (5(a), em *data-test1* e 5(b), *data-test2*). O primeiro valor corresponde novamente aos resultados obtidos por GA guiado unicamente por ϵ , para completar o estudo comparativo.

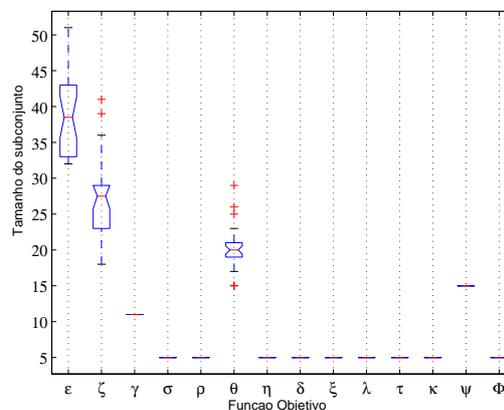


Figura 4. Tamanho dos subconjuntos de classificadores encontrados através da utilização de 13 diferentes medidas combinadas com o tamanho das soluções ζ em pares de funções objetivo usadas por NSGA-II.

Com base nesses resultados, pode-se observar que:

1. A combinação de medidas de diversidade com ζ não ajuda a encontrar subconjuntos de classificadores com elevada taxa de acerto.
2. Há de fato classificadores redundantes. A taxa de acerto do conjunto inicial de classificadores, isto é, do conjunto de 100 kNN é 96.28% em *data-test1*. Esse resultado é 0.07% pior do que a taxa de acerto das soluções encontradas usando ζ e ϵ como par de função objetivo (média de 96.35% em *data-test1* (figura 5).

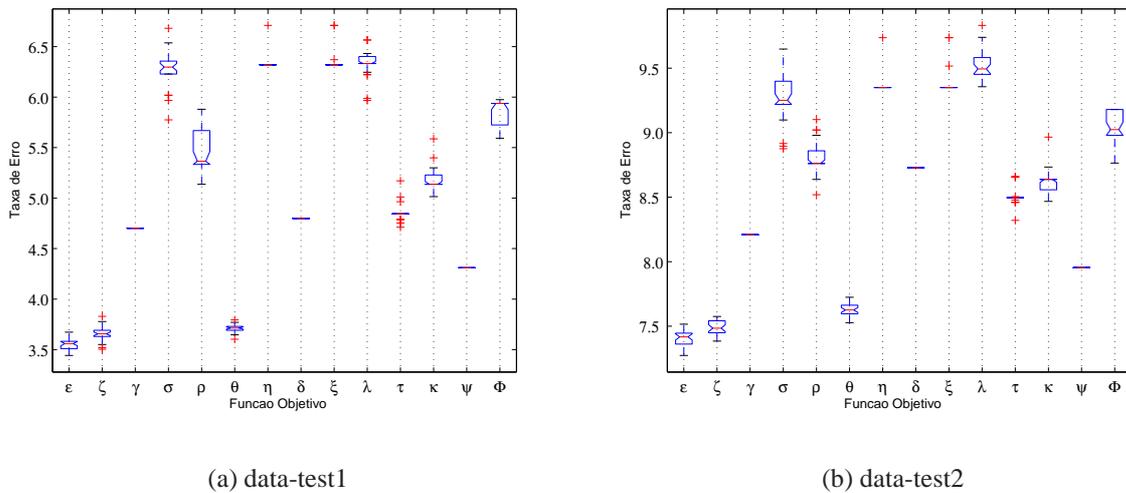


Figura 5. Desempenho dos subconjuntos de classificadores encontrados por NSGA-II guiado por pares de funções objetivo compostas por tamanho dos subconjunto ζ e 12 medidas de diversidade e taxa de acerto. O desempenho foi calculado em *data-test1* (figura 5(a)) e em *data-test2* (figura 5(b)).

Por outro lado, a média de tamanho dos subconjuntos gerados por ζ e ϵ é de 27 classificadores (figura 4).

- Os subconjuntos com maior taxa de acerto foram encontrados na otimização mono-objetiva guiada unicamente por ϵ .

5. Conclusões

Este artigo apresentou os resultados de um estudo experimental comparando critérios de busca aplicados ao problema de seleção de subconjuntos de classificadores. Os experimentos foram divididos em três séries. Na primeira série de experimentos, a taxa de erro e 12 medidas de diversidade foram comparadas diretamente em um processo de otimização mono-objetiva. Na segunda série, as 12 medidas de diversidades foram combinadas com taxa de erro para compor pares de funções objetivo em um processo de otimização multi-objetiva. Por fim, os critérios de taxa de erro e as 12 medidas de diversidade foram combinadas com tamanho dos subconjuntos de classificadores para compor pares de funções objetivo, novamente em um processo de otimização multi-objetiva. Os resultados obtidos mostram que os melhores subconjuntos de classificadores são encontrados quando taxa de erro é usada como única função objetivo. A combinação em pares de funções objetivo, tanto com diversidade quanto com tamanho dos subconjuntos de classificadores, não ajuda a melhorar a qualidade das soluções encontradas em otimização mono-objetiva. Entretanto, é importante destacar que diversidade combinada com taxa de erro supera os resultados obtidos utilizando-se diversidade como única função objetivo. Outra conclusão interessante é que a redução do número de classificadores ocorre sem que o tamanho do conjunto seja usado explicitamente como função objetivo.

Os próximos passos desta pesquisa envolverão experimentos com seleção dinâmica de classificadores.

Referências

- Aksela, M. and Laaksonen, J. (2006). Using diversity of errors for selecting members of a committee classifier. *Pattern Recognition*, 39(4):608–623.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140.
- Deb, K. (2001). *Multi-Objective Optimization using Evolutionary Algorithms*. John Wiley & Sons, LTD.
- Dietterich, T. G. (2000). Ensemble methods in machine learning. In *Proceedings of I International Workshop on Multiple Classifier System*, pages 1–15, Cagliari, Italy.
- Gother, P. (1995). *NIST Special Database 19 - Handprinted forms and characters database*. National Institute of Standard and Technology - NIST : database CD documentation.
- Ho, T. (1998). The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844.
- Kuncheva, L. and Whitaker, C. (2002). Measures of diversity in classifier ensembles an their relationship with the ensemble accuracy. *Machine Learning*, 51:181–207.
- Oliveira, L., Sabourin, R., Bortolozzi, F., and Suen, C. (2002). Automatic recognition of hand-written numerical strings: A recognition and verification strategy. *IEEE Transactions on PAMI*, 24(11):1438–1454.
- Opitz, D. W. and Shavlik, J. W. (1996). Actively searching for an effective neural-network ensemble. *Connection Science*, 8(3/4):337–353.
- Ruta, D. and Gabrys, B. (2005). Classifier selection for majority voting. *Information Fusion*, 6(1):163–168.
- Tremblay, G., Sabourin, R., and Maupin, P. (2004). Optimizing nearest neighbour in random subspaces using a multi-objective genetic algorithm. In *Proceedings of International Conference Pattern Recognition*, pages 208–211, Cambridge, UK.
- Zenobi, G. and Cunningham, P. (2001). Using diversity in preparing ensembles of classifiers based on different feature subsets to minimize generalization error. In *Proceedings of XII European Conference on Machine Learning*, pages 576–587, Freiburg, Germany.
- Zhou, Z.-H., Wu, J., and Tang, W. (2002). Ensembling neural networks: Many could be better than all. *Artificial Intelligence*, 137(1-2):239–263.