

# Um classificador *kernel* composto por um comitê de Perceptrons balanceados

Karen Braga Enes, Saulo Moraes Villela, Raul Fonseca Neto

Departamento de Ciência da Computação, Universidade Federal de Juiz de Fora, Brasil

Email: karenenes@ice.ufjf.br, {saulo.moraes, raulfonseca.neto}@ufjf.edu.br

**Resumo**—Recentemente, abordagens baseadas em comitês de classificadores e métodos *ensemble* têm sido bastante exploradas por serem uma alternativa simples e eficaz para a construção de classificadores mais acurados. A melhoria da capacidade de generalização de um *ensemble* está diretamente relacionada a acurácia de cada classificador individual, bem como a diversidade dos classificadores que o compõem. Sendo assim, contribuições nesse escopo ainda são relevantes. Nesse trabalho, é apresentada uma extensão de um modelo *ensemble* baseado em Perceptrons balanceados que permite a inclusão de funções *kernel* e a solução de problemas não-linearmente separáveis. Visando a melhoria da acurácia do classificador individual, o hiperplano solução é balanceado. Além disso, uma medida de dissimilaridade é introduzida com intuito de maximizar a diversidade do *ensemble*. Essa estratégia permite a aceitação de um novo componente no comitê se, e somente se, uma distância mínima pré-estabelecida é mantida entre o novo candidato e todos os outros componentes. Um estudo experimental foi conduzido em bases de dados não-linearmente separáveis. Os resultados obtidos mostraram que o método proposto foi capaz de superar outros algoritmos avaliados, como o AdaBoost e o SVM, na maior parte dos casos testados. Além disso, o método proposto superou consistentemente o classificador de base empregado.

**Keywords**—*Ensemble; Perceptron kernel; Classificação.*

## I. INTRODUÇÃO

Recentemente, abordagens baseadas em métodos *ensemble* têm sido amplamente difundidas, tornando-se uma área de pesquisa importante e ativa, por serem uma alternativa eficaz para a criação de classificadores mais acurados. De fato, resultados teóricos e empíricos demonstram que métodos *ensemble* apresentam desempenho superior em relação aos classificadores individuais que os compõem [1], [2], [3].

Um *ensemble* consiste em um comitê de classificadores cujas hipóteses individuais são induzidas separadamente e as decisões referentes a cada hipótese são combinadas através de um método de consenso para a classificação de novas amostras apresentadas. Entre as estratégias mais comuns para combinar decisões destacam-se a média e o voto, ponderado ou não, das soluções.

Para a construção de um modelo *ensemble*, dois aspectos fundamentais devem ser observados, a acurácia individual do classificador de base e a diversidade dos componentes do comitê [1]. Em relação a acurácia individual, um classificador é dito acurado se a taxa de erro fornecida por ele é inferior àquela fornecida por qualquer outra escolha aleatória. Além disso, a diversidade dos componentes é extremamente importante para possibilitar a correção de amostras classificadas incorretamente por um ou mais classificadores individuais.

Nesse sentido, a diversidade do comitê implica diretamente na independência de seus componentes.

Considerando as duas premissas anteriormente descritas, este trabalho apresenta uma extensão da formulação do método *Ensemble of Balanced Perceptrons* [4], para possibilitar a utilização de funções *kernel* e a solução de problemas não-linearmente separáveis, chamado *Ensemble de Perceptrons Kernel Balanceados (Ensemble of Balanced Perceptrons Kernel - EBPK)*. A escolha do modelo Perceptron *kernel* [5], como classificador de base, deve-se a simplicidade e eficácia do modelo e sua aplicabilidade em dados não-linearmente separáveis, característica frequentemente encontrada quando são tratados dados de aplicações reais. O Perceptron *kernel* balanceado é empregado com intuito de melhorar a acurácia do classificador de base. O hiperplano correspondente à solução final do Perceptron é deslocado em direção à máxima margem daquela solução. Embora a movimentação do hiperplano não altere a direção da solução final, a capacidade de generalização da solução geralmente é melhorada, já que a distância do hiperplano separador para cada uma das classes, nessa direção, é a maior possível.

Em adição, uma estratégia para maximização da diversidade dos componentes do comitê é apresentada. Essa estratégia consiste na combinação de duas heurísticas. A primeira consiste na geração de uma permutação aleatória do conjunto de treinamento, para cada membro do comitê. A alteração na ordem em que os dados de entrada são apresentados pode fornecer uma solução diferente do algoritmo, já que a ordem de correção das amostras é alterada. No entanto, apenas essa heurística é insuficiente para geração de um *ensemble* completamente diversificado. Por essa razão, uma segunda heurística foi combinada, a medida de dissimilaridade. Essa medida elimina os hiperplanos que não obedecem a uma distância mínima pré-definida em relação aos demais. Assim, apenas aqueles hiperplanos que mantiverem uma distância suficiente de cada um dos outros componentes são considerados para a construção do modelo, maximizando, portanto, a diversidade do comitê gerado. Em relação a estratégia de combinação das decisões dos classificadores individuais, dois esquemas para combinar as saídas são adotados, a média e o voto majoritário não ponderado dos componentes.

O estudo experimental apresentado nesse trabalho foi centrado em problemas não-linearmente separáveis e dividido em dois grupos de experimentos. Todos os métodos empregados na avaliação do estudo experimental fazem uso de funções *kernel* na sua implementação. Em relação ao primeiro grupo de experimentos, comparou-se o método EBPK, representado pela média das hipóteses, com o modelo *Support Vector Machines*

(SVM) [6]. Considerando o segundo grupo de experimentos, o EBPK, representado pelo voto majoritário não ponderado das hipóteses, foi comparado ao AdaBoost [7] usando um Perceptron *kernel* como classificador de base. Em todos os grupos de experimentos, resultados referentes ao modelo Perceptron *kernel* balanceado também são apresentados. Os resultados obtidos mostraram que o algoritmo proposto superou o classificador de base. Além disso, foram obtidos resultados bastante satisfatórios em relação aos outros algoritmos usados na avaliação. O método apresentado foi capaz de superar o SVM e o AdaBoost na maior parte das bases de dados testadas.

## II. TRABALHOS RELACIONADOS

A maior parte dos trabalhos desenvolvidos na área de aprendizado de máquinas, para métodos de classificação, consiste no desenvolvimento de novas estratégias com intuito de aumentar a capacidade de generalização dos modelos. Estudos atuais têm demonstrado que técnicas que envolvem a combinação de alguns classificadores de base em um único modelo, geralmente chamado de *ensemble*, tendem a prover ganhos de acurácia de classificação em relação a cada membro individual, evitando *overfitting* e reduzindo o viés [1], [8].

É possível encontrar, na literatura, resultados teóricos e empíricos em relação a métodos *ensemble* [1], [3]. Em geral, os métodos *ensemble* diferem entre si em relação a três aspectos principais: a escolha do classificador de base, a estratégia de combinação das saídas e o tratamento dos dados de entrada. Em relação ao primeiro aspecto, modelos de Árvore de Decisão e Redes Neurais Artificiais são os mais comumente empregados [1], [9], [10]. O segundo aspecto mencionado trata das estratégias usuais para combinar as saídas dos classificadores de base que incluem a média das hipóteses e também votos, ponderados ou não. Considerando o terceiro aspecto, geralmente abordagens baseadas na manipulação dos dados de treinamento para geração de múltiplas hipóteses são mais utilizadas, como ocorre em métodos como Bagging [11] e AdaBoost.

Uma característica interessante dos métodos Bagging e AdaBoost, vale ressaltar, é que caso o classificador de base escolhido seja instável, i.e., pequenas alterações no conjunto de treinamento produzem grandes modificações na hipótese gerada, então o *ensemble* gerado será diverso [11], [12]. Como, em geral, os métodos citados empregam um subconjunto diferente da base de treinamento para cada classificador individual, normalmente, o *ensemble* gerado é diverso. Enquanto o método de Bagging emprega, como saída final da classificação, um voto simples não ponderado, o AdaBoost utiliza um esquema de voto ponderado por uma função da acurácia da fase de treinamento. Estudos anteriores têm relatado que, na maior parte dos casos, o AdaBoost supera consistentemente o modelo de Bagging [12], [13].

Nos últimos anos, muito esforço tem sido empregado com intuito de construir estratégias capazes de aumentar a diversidade dos componentes dos *ensembles*. Alguns métodos focam em produzir diversidade ao aplicar estratégias de seleção de características [14], contudo abordagens baseadas em agrupamento [9] e estratégias evolucionárias [10] são as mais comumente usadas. Entretanto, não existem estudos conclusivos definindo qual medida de diversidade é a mais

adequada. Assim, contribuições com intuito de prover um aumento na diversidade dos componentes do classificador são ainda relevantes. Além disso, considerar a diversidade de um *ensemble* só faz sentido caso cada classificador individual seja suficientemente acurado. De outra forma, a diversidade não é garantia de bons resultados. Nesse sentido, alguns trabalhos propõem a utilização de classificadores não-lineares com objetivo de melhorar a acurácia do classificador de base. Em [15], um *ensemble* de SVMs não lineares é proposto no qual cada classificador de base é responsável por uma subamostragem da base de treinamento. Em [16], um modelo *ensemble* de rede neurais com *kernel* sugere que a estrutura das redes deve ser distinta para geração de componentes diversos.

O trabalho proposto nesse artigo considera contribuições relacionadas a essas duas questões. Como classificador de base, empregou-se o Perceptron *kernel* balanceado, um modelo não-linear baseado no deslocamento da solução final do modelo Perceptron *kernel*, como estratégia para melhoria da acurácia individual. Em relação a diversidade, as amostras de treinamento são manipuladas. Diferentemente do Bagging e AdaBoost, essa manipulação ocorre na base completa gerando uma permutação diferente de todo o conjunto de dados de treinamento para cada componente gerado. Finalmente, uma medida de dissimilaridade é introduzida para maximizar a distância entre os componentes do *ensemble*, permitindo a aceitação de um novo candidato se, e somente se, esse está a uma distância mínima pré-determinada de todos os outros componentes.

## III. ENSEMBLE DE PERCEPTRONS *Kernel* BALANCEADOS

Essa seção contém a descrição do método *Ensemble* de Perceptrons *Kernel* Balanceados (Ensemble of Balanced Perceptrons *Kernel* – EBPK). O modelo Perceptron *kernel* em variáveis duais é revisado e, a seguir, define-se a estratégia empregada para melhoria da acurácia do classificador de base, o modelo do Perceptron *kernel* balanceado. O modelo *ensemble* empregado consiste na combinação desse classificador de base. Posteriormente, são descritas as estratégias empregadas para tratamento dos dados de entrada com o objetivo de gerar aleatoriedade no modelo, bem como a medida de dissimilaridade para aumento da diversidade do comitê.

### A. O algoritmo Perceptron em variáveis duais

Inicialmente, considere o paradigma de aprendizado supervisionado. Uma tarefa de classificação binária pode ser definida da seguinte forma: tome o conjunto dos dados de entrada, denotado por  $Z = \{(x_i, y_i) : i \in \{1, 2, \dots, m\}\}$ , como o conjunto de treinamento de  $m$  amostras de uma função desconhecida  $f$ . Cada componente dos vetores de entrada,  $x_i \in \mathbb{R}^d$ , é chamado de atributo ou característica e pode admitir valores reais ou discretos. Os valores de  $y$  são mapeados em um conjunto discreto de classes,  $y_i \in \{-1, +1\}$ . Dado um conjunto de treinamento, um algoritmo de aprendizado é capaz de gerar um classificador dado por uma hipótese (representada por um hiperplano) em relação a função  $f : \mathbb{R}^d \rightarrow \{-1, +1\}$ . O classificador prediz os valores correspondentes de  $y$  à medida que amostras adicionais  $x$  são apresentadas. Essas amostras adicionais constituem o chamado conjunto de teste.

Para problemas não-linearmente separáveis, faz-se necessário o emprego de técnicas auxiliares capazes de

lidar com a essa característica. Uma abordagem eficaz para solucionar tal contratempo consiste no emprego de funções *kernel* na construção do classificador. Nesse sentido, torna-se necessário o mapeamento dos dados para um espaço de mais alta dimensão, chamado espaço de características, ou  $\phi$ -space, comumente representado por  $F$ . Com o mapeamento  $\phi : \mathbb{R}^d \times \mathbb{R}^d \rightarrow F$ , é possível a representação do conjunto de amostras em um espaço de mais alta dimensão,  $x \rightarrow \phi(x)$ , no qual o problema se torna linearmente separável.

A versão dual do modelo Perceptron permite a introdução de funções *kernel* e posterior solução do problema relacionado à construção de uma hipótese linear no espaço de variáveis duais. Para isso, algumas alterações no modelo proposto por Rosenblatt [17] devem ser consideradas. O vetor de pesos  $w$  é representado como um combinação linear dos vetores de entrada  $(x_i, y_i)$  da seguinte forma

$$w = \sum_{j=1}^m \alpha_j y_j \phi(x_j), \quad (1)$$

na qual  $\alpha \in \mathbb{R}^m$ ,  $\alpha \geq 0$ , é o vetor de multiplicadores associado ao conjunto de entrada.

A grande vantagem de um classificador *kernel* é que não é necessário conhecer o tipo de mapeamento ou a função  $\phi$  explicitamente. Para tanto, utiliza-se uma função *kernel*, simétrica e semi-definida positiva, definida como  $K : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ .

A formulação dual do modelo Perceptron consiste em encontrar um hiperplano separador, dado pela solução do sistema de inequações lineares, definido da seguinte forma

$$f(x_i) = \begin{cases} +1, & \sum_{j=1}^m \alpha_j y_j K(x_j, x_i) + b \geq 0 \\ -1, & \sum_{j=1}^m \alpha_j y_j K(x_j, x_i) + b < 0, \end{cases} \quad (2)$$

na qual  $b$  é o valor de *bias*.

No decorrer deste trabalho, utilizou-se um *kernel* gaussiano definido da seguinte forma

$$K(x_j, x_i) = \exp\left(-\frac{\|x_j - x_i\|_2^2}{2\sigma^2}\right), \quad (3)$$

onde  $\sigma$  é a largura da gaussiana e  $\|\cdot\|_2^2$  a norma Euclidiana ao quadrado.

Uma amostra de treinamento  $(x_i, y_i)$  é classificada incorretamente se  $y_i \left( \sum_{j=1}^m \alpha_j y_j K(x_j, x_i) + b \right) < 0$ . Caso isso ocorra, a regra de correção do modelo deve ser aplicada até que não haja mais instâncias incorretamente classificadas. A regra de correção é definida da seguinte forma

$$\begin{aligned} w &= \sum_{j=1}^m \alpha_j y_j x_j + \eta y_i x_i \\ \alpha_i^{t+1} &\leftarrow \alpha_i^t + \eta \\ b^{t+1} &\leftarrow b^t + \Delta \alpha_i y_i, \end{aligned} \quad (4)$$

na qual  $\Delta \alpha_i y_i$  refere-se ao somatório das correções nos valores dos multiplicadores considerando o respectivo sinal das classes,  $\eta$  é a taxa de aprendizagem e  $t$  a variável de iteração. Ao passo que todas as instâncias de treinamento são

corretamente classificadas, está definido o hiperplano separador solução do problema.

### B. Melhorando a acurácia do classificador individual - Perceptron Kernel Balanceado

O modelo original do algoritmo Perceptron *kernel*, devido a sua formulação, tende a manter uma distância pequena entre o hiperplano separador e as últimas amostras corrigidas. Esse fato pode culminar em um hiperplano solução desbalanceado. Em alguns casos, o valor de margem pode ser significativamente melhorado através de um reposicionamento do hiperplano, mantendo a mesma direção da solução. Dessa forma, o deslocamento no valor de *bias* movimenta o hiperplano para uma posição equidistante das duas classes. Essa nova posição é a solução de máxima margem da hipótese em questão, provavelmente implicando em uma capacidade de generalização mais elevada para o modelo.

Dados  $Z^+ = \{(x_i, y_i) \in Z : y_i = +1\}$ ,  $Z^- = \{(x_i, y_i) \in Z : y_i = -1\}$  e considerando a distância entre o hiperplano separador e as duas classes do problema, as semi-margens são definidas da seguinte forma

$$\begin{aligned} \gamma^+ &= \min \left\{ y_i \left( \sum_{j=1}^m \alpha_j y_j K(x_j, x_i) + b \right), \forall x_i \in Z^+ \right\} \\ \gamma^- &= \min \left\{ y_i \left( \sum_{j=1}^m \alpha_j y_j K(x_j, x_i) + b \right), \forall x_i \in Z^- \right\}. \end{aligned} \quad (5)$$

O deslocamento é determinado com base na média das margens. Assim, define-se o deslocamento como  $\Gamma = (\gamma^+ + \gamma^-)/2$ . Dado o valor de  $\Gamma$ , o novo valor do *bias* é dado por

$$b_{new} = b_{old} - \gamma^- + \Gamma. \quad (6)$$

Uma vez que o vetor de pesos não é alterado, a solução final do Perceptron *Kernel* Balanceado é equivalente a solução do Perceptron *kernel*. Entretanto, ao introduzir o deslocamento no *bias*, o novo hiperplano obtido mantém a margem mais larga possível, considerando a mesma solução.

### C. Introduzindo diversidade no ensemble

1) *Permutação aleatória dos dados de entrada*: Uma forma de gerar a diversidade no comitê consiste na manipulação dos dados de treinamento para a geração de múltiplas hipóteses. Para cada componente do *ensemble*, o algoritmo de aprendizado é executado com uma permutação diferente do conjunto de treinamento. Essa estratégia eleva a diversidade do *ensemble*, já que alterar a ordem de correção das amostras pode culminar em uma solução final diferente.

2) *Medida de dissimilaridade*: Para algoritmos estáveis, como o Perceptron *kernel*, simplesmente tornar aleatória a ordem das amostras de entrada não é suficiente para produzir uma diversidade de componentes adequada. Nesse caso, o *ensemble* gerado pode contar com muitos componentes similares.

Com intuito de gerar um conjunto de hipóteses o mais diverso possível, uma outra estratégia é adotada. Em adição às estratégias descritas anteriormente, define-se um valor,  $\varepsilon > 0$ , que representa a distância mínima que todos os componentes pertencentes ao comitê devem manter entre si. Em outras palavras, dado o conjunto de hipóteses  $\{h_1, h_2, \dots, h_j\}$  já

aceitas no comitê, uma nova hipótese  $h_j$  é tomada como um componente válido se, e somente se

$$\min_{i \in \{1, \dots, j\}} \{\|h_i, h_k\|_2\} > \varepsilon. \quad (7)$$

Essa estratégia é chamada de *medida de dissimilaridade*.

A medida de dissimilaridade adotada no modelo EBPK é a distância de Tanimoto [18], normalizada dos vetores normais ao espaço de características. Sejam dois vetores  $\alpha_i, \alpha_j \in F$  referentes a duas hipóteses  $h_i, h_j$ , define-se a distância entre dos mesmos da seguinte forma

$$\text{dist}(h_i, h_j) = \text{dist}(\alpha_i, \alpha_j) = 1 - \frac{\alpha_i \alpha_j}{\alpha_i^2 + \alpha_j^2 - \alpha_i \alpha_j}. \quad (8)$$

Vale ressaltar que, essa medida gera um coeficiente de dissimilaridade entre 0 e 1, no qual 0 representa vetores idênticos, cuja distância entre os mesmos é nula.

#### IV. ANÁLISE EXPERIMENTAL E DISCUSSÕES

Essa seção apresenta o estudo experimental realizado para avaliar o método proposto. O estudo foi feito em 5 bases de dados não-linearmente separáveis e, ainda, a base Sonar que, embora seja linearmente separável, dificilmente é resolvida por modelos lineares. As informações principais encontram-se na Tabela I. Todas as bases utilizadas nesse artigo estão contidas no repositório de aprendizado de máquinas da UCI [19].

A avaliação foi baseada na comparação do método proposto considerando duas estratégias de combinação de saídas, a média das hipóteses e o voto não ponderado. Com intuito de avaliar a efetividade da medida de dissimilaridade, foram considerados duas variações do método: a primeira EBPK, na qual a medida de dissimilaridade não é empregada, e a segunda EBPKd, chamada de *Ensemble* de Perceptrons *Kernel* Dissimilares Balanceados, o qual emprega a medida de dissimilaridade.

Os experimentos são iniciados com dois problemas simples para observar a influência da medida de dissimilaridade. Em seguida, o estudo experimental é dividido em dois grupos. O primeiro grupo de experimentos considera métodos que podem ser representados por uma única hipótese ou a média de hipóteses. Nesse grupo, são apresentados os resultados para o m-EBPK e o m-EBPKd, nos quais m- indica que a hipótese final é obtida a partir da média das hipóteses do comitê. Esses resultados são comparados a um classificador SVM. O segundo grupo trata dos métodos cuja solução é obtida por esquemas de votação. Os resultados são apresentados para os métodos v-EBPK e v-EBPKd e comparados ao modelo AdaBoost. Essa separação é considerada para promover justiça em relação as saídas dos métodos.

Em ambos os grupos, são apresentados resultados referentes ao modelo individual do Perceptron *Kernel* Balanceado (PB). Ao fazê-lo, o objetivo é validar a premissa básica da construção de *ensembles*, que diz que um modelo *ensemble* deve obter uma performance superior ao classificador individual em questão. As implementações dos algoritmos SVM e AdaBoost consideradas nesse trabalho referem-se ao *Sequential Minimal Optimization* [20] sem flexibilização de margem e AdaBoost.M1 [7] com um Perceptron *kernel* como classificador de base, respectivamente.

TABELA I. INFORMAÇÕES SOBRE AS BASES DE DADOS CONSIDERADAS.

Base	Atributos	Amostras		
		Pos.	Neg.	Total
Sonar	60	97	111	208
Ionosphere	34	225	126	351
Tic Tac	9	626	332	958
Bupa	6	145	200	345
Pima	8	268	500	768
Wine	13	107	71	178

##### A. Avaliação da medida de dissimilaridade

Objetivando investigar a influência da medida de dissimilaridade, testes foram realizados em dois problemas simples. Ambos os problemas são linearmente separáveis e bidimensionais com 8 amostras de treinamento para construção do comitê e foram escolhidos para facilitar a visualização dos resultados, já que o cômputo da medida de dissimilaridade deve ser feito no espaço de características, no qual o problema analisado passa a ser linearmente separável. Os métodos investigados consideram o EBPK sem balanceamento e sem medida de dissimilaridade para comparação com o EBPKd empregando um *kernel* linear. A Fig. 1 apresenta os resultados.

Em relação ao primeiro problema, Fig. 1-(a),(b), os comitês gerados têm 5 componentes. Observa-se que o modelo EBPK sem balanceamento e sem medida de dissimilaridade, Fig. 1-(a), apesar de considerar em sua formulação a permutação aleatória dos dados de entrada, produz componentes muito similares, ou seja, pouco diversificados. Por outro lado, o modelo EBPKd, Fig. 1-(b), conta com componentes capazes de preencher melhor o espaço disponível com apenas 5 componentes, tornando assim, o *ensemble* gerado bastante diversificado. Considerando o segundo problema, Fig. 1-(c),(d), foram gerados comitês com 10 componentes. De maneira análoga, é possível observar que os componentes do comitê EBPK sem balanceamento e sem medida de dissimilaridade, Fig. 1-(c), são pouco diversificados e não preenchem todo o espaço. Da mesma forma, ao introduzir a medida de dissimilaridade e o balanceamento no modelo, Fig. 1-(d), o comitê gerado ficou diverso e bem distribuído em relação ao espaço em questão. Dessa forma, estima-se que a medida de dissimilaridade é efetiva e capaz de maximizar a diversidade no *ensemble*.

##### B. Definição de parâmetros

Considerando a natureza aleatória dos modelos abordados, em todos os experimentos foi empregada uma estratégia de *cross-validation* em 10x10-10-fold, exceto para o caso do SVM, em que foi empregado a estratégia em 1x10-10-fold. Esse esquemas foram adotados com intuito de reduzir o vies dos métodos analisados. Em relação à estratégia de *cross-validation*, em cada *fold* sempre foi mantida a porcentagem de dados referentes a cada classe [21]. Visando comparações mais precisas, para cada base de dados, sempre foram selecionados os mesmos 10 subconjuntos do *cross-validation*, preservando a geração da semente associada ao processo aleatório. O parâmetro referente a taxa de aprendizado foi fixado em  $\eta = 0.05$ , para todos os Perceptrons *kernel*. Utilizou-se um *kernel* gaussiano com largura  $\sigma = 1.0$  em todos os modelos avaliados. A efetividade da acurácia de classificação dos modelos analisados foi avaliada por meio da taxa de

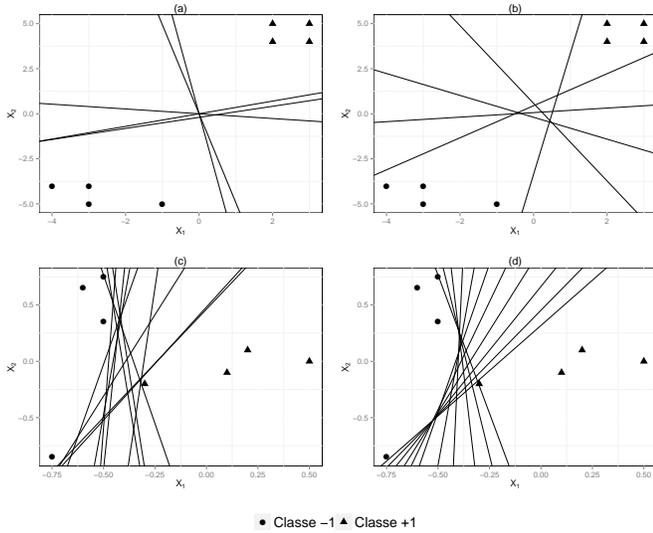


Fig. 1. Medida de dissimilaridade: (a)(c) *Ensemble* obtido a partir do método EBPK sem balanceamento e sem dissimilaridade; (b)(d) *Ensemble* obtido a partir do método EBPKd.

erro convencional, i.e., a porcentagem de amostras de teste classificadas incorretamente, considerando a média dos erros obtidos em cada execução.

1) *Análise da influência na variação do tamanho do comitê*: Com intuito de investigar o efeito causado pela variação do tamanho do *ensemble* na capacidade de generalização do modelo obtido, foram realizados testes nas referidas bases de dados. Executou-se um esquema de *cross-validation* segundo a configuração  $10 \times 10$ -*fold* considerando as duas formas de combinação de saída empregadas nesse estudo e aplicada ao método EBPKd. O tamanho do *ensemble* foi variado segundo  $s = \{1, \dots, 10, 20, \dots, 100, 200, \dots, 1000\}$ . Entretanto, optou-se por condensar os resultados apresentando apenas três valores tendo em vista a dificuldade de visualização e avaliação.

TABELA II. RESULTADOS DA COMPARAÇÃO ENTRE DIFERENTES TAMANHOS DE COMITÊ CONSIDERANDO O ERRO MÉDIO DE CLASSIFICAÇÃO.

Combinação	Tamanho	Ionosphere	Bupa
Média	10	<b>5,92 ± 0,51</b>	<b>37,73 ± 0,68</b>
	100	5,96 ± 0,53	39,79 ± 0,67
	1000	5,99 ± 0,47	40,45 ± 0,27
Votação Majoritária	10	<b>6,09 ± 0,43</b>	37,55 ± 0,74
	100	6,10 ± 0,42	37,51 ± 0,64
	1000	6,11 ± 0,31	<b>37,36 ± 0,70</b>

Considerando todos os resultados obtidos, observou-se que, na maioria dos casos, o aumento do número de componentes do comitê não necessariamente resulta em redução da taxa de erro. Além disso, em alguns casos, observa-se que, quanto maior a quantidade de componentes no comitê, maior o erro de generalização obtido. Esse efeito pode ser observado claramente na Tabela II. Tais resultados sugerem que o aumento no tamanho do *ensemble* pode culminar em *overfitting* do modelo, como mencionado em [13]. Como consequência, a geração de um número de componentes pequenos tende a evitar *overfitting* no comitê. Por essa razão, sem eventuais perdas significativas da capacidade de generalização, optou-se por fixar o tamanho

do modelo *ensemble* em  $10 + 1$  componentes, com intuito de evitar possíveis empates ao utilizar a estratégia de votação.

2) *Medida de dissimilaridade*: A medida de dissimilaridade é o único parâmetro não fixado no modelo tratado. Esse parâmetro é responsável por imprimir diversidade na construção do *ensemble*. A escolha de um valor adequado depende tanto da complexidade do problema, quanto do tamanho pretendido de comitê. Vale ressaltar que, como a medida de dissimilaridade considera o cômputo da distância de Tanimoto, quanto maior o valor determinado, maior a diferença de distância estabelecida para as hipóteses geradas. Nesse sentido, para a construção de comitês com muitos componentes, valores pequenos devem ser estabelecidos para garantir admissão dos componentes requisitados para construção do *ensemble*.

No decorrer do estudo experimental, a medida de dissimilaridade foi definida com base em testes empíricos para cada modelo analisado. Foi definido um valor de forma que o número de hipóteses rejeitadas fosse três vezes maior que o número de componentes no *ensemble*.

### C. Resultados Numéricos

1) *Ensemble representado pela média das hipóteses*: Essa seção tem por objetivo discutir os resultados obtidos pelo método proposto, combinado pela média dos componentes do *ensemble* (m-EBPK e m-EBPKd), com os modelos SVM e PB. Todos os métodos foram aplicados às bases de dados descritas na Seção IV-B levando em consideração os parâmetros definidos na Seção IV-C. A Tabela III apresenta os resultados de erro médio obtido para cada classificador.

TABELA III. COMPARAÇÃO ENTRE OS MÉTODOS ENSEMBLE COMBINADOS PELA MÉDIA E O SVM.

Base	PB	m-EBPK	m-EBPKd	SVM
Sonar	15,67 ± 1,32	13,78 ± 1,09	14,26 ± 1,05	<b>13,40 ± 0,77</b>
Ionosphere	6,50 ± 0,80	<b>5,71 ± 0,66</b>	5,92 ± 0,51	7,15 ± 0,40
Tic Tac	2,18 ± 0,46	0,05 ± 0,07	<b>0,02 ± 0,04</b>	1,68 ± 0,03
Bupa	40,59 ± 2,19	38,77 ± 0,76	<b>37,73 ± 0,68</b>	40,45 ± 0,27
Pima	35,41 ± 1,48	<b>34,89 ± 0,01</b>	<b>34,89 ± 0,01</b>	<b>34,89 ± 0,00</b>
Wine	36,38 ± 1,33	39,12 ± 0,95	<b>36,30 ± 1,18</b>	39,90 ± 0,00

Através dos resultados apresentados, observa-se que as abordagens m-EBPK e m-EBPKd superam o classificador individual PB em todos os casos avaliados. O método proposto apresenta também uma taxa de erro de generalização inferior ao modelo SVM em 5 das 6 bases de dados testadas. Além disso, é possível observar ainda que a inclusão da medida de dissimilaridade, como esperado, produz efeitos positivos na redução da taxa de erro do modelo, já que ocorre um aumento na diversidade dos componentes. Assim, a média das hipóteses obtida através do m-EBPKd é mais representativa do que a média obtida pelo m-EBPK.

2) *Ensemble representado pelo voto das hipóteses*: A segunda parte da análise experimental compara o método proposto, combinado pelo voto majoritário não ponderado dos componentes (v-EBPK e v-EBPKd), aos modelos PB e AdaBoost. O AdaBoost utiliza um Perceptron *kernel* simples como classificador de base e considera o voto ponderado dos membros do comitê. A Tabela IV apresenta os resultados de erro médio obtidos para os problemas em questão.

Observa-se, a partir dos resultados reportados, que as formulações v-EBPK e v-EBPKd apresentam resultados bastante

TABELA IV. COMPARAÇÃO ENTRE OS MÉTODOS ENSEMBLE COMBINADOS PELO VOTO MAJORITÁRIO E O ADABOOST.

Bases	PB	v-EBPK	v-EBPKd	AdaBoost
Sonar	15,67 ± 1,32	<b>13,87 ± 1,00</b>	14,56 ± 1,36	14,99 ± 1,58
Ionosphere	6,50 ± 0,80	<b>5,86 ± 0,58</b>	6,09 ± 0,43	12,10 ± 0,92
Tic Tac	2,18 ± 0,46	<b>0,10 ± 0,09</b>	0,11 ± 0,10	0,63 ± 0,24
Bupa	40,59 ± 2,19	<b>37,29 ± 0,64</b>	37,55 ± 0,74	41,62 ± 1,98
Pima	35,41 ± 1,48	<b>34,68 ± 0,49</b>	34,70 ± 0,62	36,51 ± 0,97
Wine	36,38 ± 1,33	33,82 ± 1,18	32,93 ± 1,38	<b>28,24 ± 1,53</b>

satisfatórios. Ambas as abordagens superam o classificador individual PB. Além disso, o método proposto apresenta uma taxa de erro inferior ao AdaBoost em 5 das 6 bases consideradas nesse estudo. Em relação a eficácia da medida de dissimilaridade é possível observar que o emprego da medida não gerou redução na taxa de erro do modelo. Isso aconteceu devido ao esquema de combinação das saídas, no qual o voto não ponderado garante o mesmo peso na votação a todos os componentes do comitê.

## V. CONCLUSÕES E TRABALHOS FUTUROS

Nesse trabalho foi proposta uma extensão do método *Ensemble of Balanced Perceptrons*, permitindo a utilização de funções *kernel* e a solução de problemas não-linearmente separáveis. O classificador de base consiste em uma modificação do modelo Perceptron *kernel*, com intuito de aumentar a acurácia do mesmo. Além disso, uma medida de dissimilaridade foi aplicada como uma alternativa para elevar a diversidade do comitê gerado. O modelo *ensemble* tratado combina um conjunto de Perceptrons *kernel* balanceados selecionados por meio da medida de dissimilaridade. O estudo experimental realizado mostrou que os resultados obtidos são relevantes. Comprovou-se a eficácia da medida de dissimilaridade quando a estratégia de combinação adotada é a média dos classificadores. Mostrou-se, ainda, que essa medida não é eficiente, como esperado, quando utiliza-se o voto majoritário não ponderado das amostras. Além disso, confirma-se a premissa de desenvolvimento de *ensembles* ao mostrar que o modelo estudado foi superior, em acurácia, ao modelo individual PB. Em relação ao método avaliado combinado pela média das hipóteses, observou-se que o EBPK foi superior, em acurácia de classificação, ao SVM na maior parte dos casos. Considerando o método avaliado combinado pela estratégia de votação majoritária, os resultados obtidos mostraram que o EBPK é capaz de superar o AdaBoost na maioria das bases de dados avaliadas.

## AGRADECIMENTOS

Os autores agradecem à CAPES pelo apoio financeiro.

## REFERÊNCIAS

- [1] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple classifier systems*. Springer, 2000, pp. 1–15.
- [2] K. Tumer and J. Ghosh, "Analysis of decision boundaries in linearly combined neural classifiers," *Pattern Recognition*, vol. 29, no. 2, pp. 341–348, 1996.
- [3] L. I. Kuncheva, *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, 2004.
- [4] K. B. Enes, S. M. Villela, and R. Fonseca Neto, "A novel ensemble approach based on balanced perceptrons applied to microarray datasets," To appear in Proceedings of the 4th Brazilian Conference on Intelligent Systems, 2015, p. to appear.

- [5] A. Aizerman, E. M. Braverman, and L. Rozoner, "Theoretical foundations of the potential function method in pattern recognition learning," *Automation and remote control*, vol. 25, pp. 821–837, 1964.
- [6] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [7] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *Proceedings of the Thirteenth International Conference on Machine Learning*. Morgan Kaufmann, 1996.
- [8] L. Breiman, "Arcing classifier," *The annals of statistics*, vol. 26, no. 3, pp. 801–849, 1998.
- [9] H. Parvin, M. MirnabiBaboli, and H. Alinejad-Rokny, "Proposing a classifier ensemble framework based on classifier selection and decision tree," *Engineering Applications of Artificial Intelligence*, vol. 37, pp. 34–42, 2015.
- [10] J. Tian, M. Li, F. Chen, and J. Kou, "Coevolutionary learning of neural network ensemble for complex classification tasks," *Pattern Recognition*, vol. 45, no. 4, pp. 1373–1385, 2012.
- [11] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.
- [12] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization," *Machine learning*, vol. 40, no. 2, pp. 139–157, 2000.
- [13] J. R. Quinlan, "Bagging, boosting, and c4.5," in *AAAI/IAAI, Vol. 1*, 1996, pp. 725–730.
- [14] P. Cunningham and J. Carney, "Diversity versus quality in classification ensembles based on feature selection," in *Machine Learning: ECML 2000*. Springer, 2000, pp. 109–116.
- [15] W.-C. Lai, P.-H. Huang, Y.-J. Lee, and A. Chiang, "A distributed ensemble scheme for nonlinear support vector machine," in *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2015 IEEE Tenth International Conference on*. IEEE, 2015, pp. 1–6.
- [16] A. Adamu, T. Maul, A. Bargiela, and C. Roadknight, "Preliminary experiments with ensembles of neurally diverse artificial neural networks for pattern recognition," in *Recent Advances in Information and Communication Technology 2015*. Springer, 2015, pp. 85–96.
- [17] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, pp. 386–408, 1958.
- [18] S. E. Umbaugh, *Computer imaging: digital image analysis and processing*. CRC press, 2005.
- [19] K. Bache and M. Lichman, "Uci machine learning repository," 2013. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [20] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," 1998.
- [21] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2*. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1995, pp. 1137–1143.