

Combinação de Classificadores Simbólicos Utilizando Medidas de Avaliação de Regras e Algoritmos Genéticos

Flávia Cristina Bernardini and Maria Carolina Monard

Departamento de Ciência da Computação - SCC

ICMC/USP São Carlos

Tese de doutorado disponível em:

<http://www.teses.usp.br/teses/disponiveis/55/55134/tde-29092006-110806/>

Data da conclusão: 29 de agosto de 2006

flavia@addlabs.uff.br, mcmonard@icmc.usp.br

Abstract. A qualidade das hipóteses induzidas pelos atuais sistemas de aprendizado de máquina supervisionado depende da quantidade dos exemplos no conjunto de treinamento. Por outro lado, muitos dos sistemas de aprendizado de máquina conhecidos não estão preparados para trabalhar com uma grande quantidade de exemplos. Entretanto, grandes conjuntos de dados são típicos em mineração de dados. Uma maneira para resolver esse problema consiste em construir *ensembles* de classificadores, e outra maneira é utilizar algoritmos genéticos para combinar os classificadores construídos em um único classificador final. Nesta tese de doutorado, foram propostos métodos para construção de *ensembles* de classificadores e foi proposto um algoritmo genético para combinação de classificadores visando resolver os problemas relacionados a extração de conhecimento de grandes bases de dados utilizando algoritmos de aprendizado de máquina.

1 Introdução

Neste artigo, são descritas, de forma sucinta, as principais contribuições da tese de doutorado de Flávia Cristina Bernardini [7]¹ O tema central da tese de doutorado é a utilização de Aprendizado de Máquina (AM) simbólico na etapa de extração de conhecimento, no processo de Mineração de Dados, tendo como foco principal a combinação de classificadores simbólicos. Este artigo está organizado como segue: Na Seção 2, são descritas as principais contribuições da tese de doutorado; na Seção 3 são descritos alguns conceitos de AM simbólico; na Seção 4 são descritos os métodos de construção de *ensembles* de classificadores propostos na tese de doutorado; na Seção 5 é descrita a interação entre Algoritmos Genéticos (AGs) e aprendizado de máquina simbólico e o AG proposto;

¹ Por restrições de espaço, somente são referenciados os trabalhos oriundos da tese de doutorado — <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-29092006-110806/>. As demais referências relacionadas aos temas aqui tratados, bem como a sua revisão bibliográfica, encontram-se na tese do doutorado.

na Seção 6 são descritos os resultados obtidos nos experimentos realizados; e na Seção 7 são apresentadas as conclusões do trabalho de doutorado realizado.

2 Contribuições da Tese

As seguintes duas questões foram colocadas para serem respondidas nesta tese:

1. É possível obter um bom poder de predição combinando classificadores simbólicos induzidos sobre conjuntos dispersos de dados sobre algum domínio em questão? Ou seja, dado um conjunto de dados grande o suficiente a ponto de não poder ser manipulado pelos algoritmos de aprendizado simbólico disponíveis atualmente, é possível extrair amostras (disjuntas ou não) desse conjunto de dados, induzir classificadores, combiná-los e melhorar o poder de predição desses classificadores, com a capacidade de também fornecer uma explicação relacionada à classificação de novos exemplos?
2. Mesmo em conjuntos tratáveis por algoritmos de aprendizado simbólico, induzindo classificadores com diferentes algoritmos sobre a mesma base de dados, é possível melhorar a predição desses classificadores, combinando-os com os métodos propostos, mantendo a capacidade de explicação?

Os objetivos deste trabalho, para responder a essas questões, utilizando algoritmos de aprendizado de máquina simbólicos que expressam o conhecimento como regras proposicionais *if-then*, compreendem:

1. Propor e avaliar métodos para construção de *ensembles* de classificadores simbólicos que possam fornecer uma explicação de suas decisões; e
2. Propor um algoritmo genético para evoluir classificadores simbólicos em um único classificador simbólico, preferencialmente mais preciso que os classificadores individuais.

Tendo em vista os objetivos acima descritos, as principais contribuições desta tese de doutoramento podem ser reunidas nos seguintes três principais grupos:

1. Novos métodos de construção de *ensembles* de classificadores simbólicos que explicam suas decisões;
2. Um método para resumir explicações fornecidas pelos *ensembles* construídos utilizando os métodos de construção de *ensembles* propostos;
3. Um algoritmo genético para evoluir classificadores simbólicos em um único classificador, de maneira a facilitar a codificação dos indivíduos (classificadores), característica essa que facilita a proposta e utilização de diferentes funções de avaliação dos classificadores (indivíduos) que constituem a população do AG.

Essas contribuições foram divulgadas no meio científico e deram lugar a diversas publicações. A seguir, são descritas cada uma das contribuições com maiores detalhes.

3 Aprendizado de Máquina Simbólico

No problema padrão de AM supervisionado, ao algoritmo de aprendizado de máquina é dado um conjunto de exemplos de treinamento S com N exemplos $T_i, i = 1, \dots, N$, escolhidos de um domínio \mathcal{X} com uma distribuição D fixa, desconhecida e arbitrária, da forma $\{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$ para alguma função desconhecida $y = f(\mathbf{x})$. Os \mathbf{x}_i são tipicamente vetores da forma $(x_{i1}, x_{i2}, \dots, x_{iM})$, com valores discretos ou contínuos, onde x_{ij} refere-se ao valor do atributo j , denominado X_j , do exemplo T_i . Os valores y_i referem-se ao valor do atributo Y , freqüentemente denominado classe. Em problemas de classificação, tratados neste trabalho, o atributo classe y_i é discreto, ou seja, $y_i \in \{C_1, C_2, \dots, C_{N_{Cl}}\}$. A partir do conjunto de treinamento S , um *classificador* \mathbf{h} é induzido. Uma maneira de avaliar esse classificador consiste em coletar informações das decisões tomadas pelo classificador em um conjunto de teste S . Para cada par (C_i, C_j) , sendo C_i e C_j pertencente ao conjunto $\{C_1, \dots, C_{N_{Cl}}\}$, calcula-se o valor da função $M(C_i, C_j)$. Essa função $M(C_i, C_j)$ é definida pelo número de exemplos em S que são pertencentes a C_i e foram classificados por \mathbf{h} como sendo pertencentes a C_j . A partir de tais valores, diversas medidas de avaliação de hipóteses podem ser utilizadas. Duas medidas freqüentemente consideradas para avaliar a performance de uma hipótese \mathbf{h} são acurácia, F_1 , precisão e sensibilidade, definidas respectivamente como $Acc(\mathbf{h}) = \frac{\sum_{i=1}^{N_{Cl}} M(C_i, C_i)}{N}$; e $F_1(\mathbf{h}) = \frac{2 \times Prec(\mathbf{h}) \times Recall(\mathbf{h})}{Prec(\mathbf{h}) + Recall(\mathbf{h})}$, $Prec(\mathbf{h}) = \frac{\sum_{i=1}^{N_{Cl}} M(C_i, C_i)}{\sum_{i=1}^{N_{Cl}} M(C_i, C_i) + \sum_{j=1}^{N_{Cl}} M(C_i, C_j)}$ e $Recall(\mathbf{h}) = \frac{\sum_{i=1}^{N_{Cl}} M(C_i, C_i)}{\sum_{i=1}^{N_{Cl}} M(C_i, C_i) + \sum_{j=1}^{N_{Cl}} M(C_j, C_i)}$.

No caso de aprendizado simbólico proposicional, \mathbf{h} pode ser transformado em um conjunto de regras **if-then**, não ordenadas ou disjuntas, tal que $\mathbf{h} = \{R_1, R_2, \dots, R_{N_R}\}$. Uma *regra proposicional* R apresenta a forma **if** B **then** H ou, simbolicamente, $B \rightarrow H$, onde H é a *cabeça*, ou a conclusão da regra R , e B é o *corpo*, ou condição de R . H e B são ambos complexos sem atributos em comum. Um complexo é uma disjunção de conjunções de testes de atributos da forma X_i *op* *Valor*, onde X_i é o nome do atributo, *op* é um operador pertencente ao conjunto $\{=, \neq, <, \leq, >, \geq\}$ e *Valor* é um valor válido para o atributo X_i . Uma *regra de classificação* assume a forma **if** B **then** classe = C_i . Ou seja, a cabeça H de uma regra de classificação é classe = C_i , com $C_i \in \{C_1, \dots, C_{N_{Cl}}\}$. Neste trabalho, as regras que compõem \mathbf{h} são regras de classificação.

A *cobertura* de uma regra $R = B \rightarrow H$ é definida como segue: exemplos que satisfazem B (o corpo da regra) compõem o conjunto de cobertura de R ; em outras palavras, esses exemplos são cobertos por R . Dada uma regra $R = B \rightarrow H$ e um conjunto de exemplos S , uma maneira de avaliar essa regra é utilizando medidas de avaliação de regras. Para o cálculo dessas medidas, são necessárias algumas informações a respeito de cada regra: o número de exemplos em S para os quais H é verdade e B é verdade (hb); o número de exemplos em S para os quais H é falso e B é verdade ($\bar{h}b$); o número de exemplos em S para os quais H é falso e B é falso ($\bar{h}\bar{b}$); e o número de exemplos em S para os quais H é verdade e B é falso ($h\bar{b}$). Algumas medidas de avaliação de regras frequentemente consideradas são cobertura, acurácia e Laplace, definidas respectivamente como $Cov(R) = h$, $Acc(R) = \frac{hb}{b}$ e $Lacc(R) = (bh + 1)/(bh + \bar{h}b + N_{Cl})$.

Como mais de uma regra com diferentes conseqüentes podem cobrir um exemplo, um conjunto de regras pode ser utilizado de diferentes maneiras para constituir um classificador. Uma delas, denominada *Single Rule classification method (SR)*, consiste em escolher, dentre todas as regras que cobrem o exemplo, a melhor regra de acordo com um critério de avaliação de regras, como por exemplo uma medida de avaliação de regras, para classificar o exemplo. Outro método, denominado *Multiple Rule classification method (MR)*, considera a combinação de todas as regras que cobrem o exemplo. Na tese, o método foi implementado como segue: considere MR_{C_i} o conjunto de regras que cobrem o exemplo a ser classificado, cujo conseqüente é a classe C_i . Para cada MR_{C_i} , a medida de qualidade da disjunção dessas regras é computada e a instância é classificada de acordo com o melhor critério de qualidade computado entre os valores MR_{C_i} .

4 Uma Proposta para Combinar Classificadores Simbólicos

O processo de construção de um *ensemble* de classificadores pode ser dividido em duas sub-tarefas [8]: (1) indução de um conjunto de classificadores; e (2) combinação das decisões dos classificadores induzidos para classificar novos exemplos.

Na proposta da tese de doutorado, a primeira tarefa dos métodos de construção de *ensembles* propostos é realizada da maneira usual: dados o número de classificadores a serem induzidos, L , e um conjunto de dados (exemplos), S , primeiramente são retiradas L amostras S_1, \dots, S_L de S com ou sem reposição. Cada uma dessas amostras, *i.e* conjunto de exemplos de treinamento, é usada como entrada para um algoritmo de aprendizado simbólico para se induzir L hipóteses (classificadores) denominadas $\mathbf{h}_1, \dots, \mathbf{h}_L$. Após, dado um novo exemplo \mathbf{x} a ser classificado, as decisões individuais do conjunto de L hipóteses $\{\mathbf{h}_1, \dots, \mathbf{h}_L\}$ devem ser combinadas para gerar a classificação final. Para a decisão de cada hipótese \mathbf{h} , foram utilizados os métodos *SR* e *MR*, utilizando as medidas de *Acc* e *Lap* e outras duas medidas de avaliação. Para combinar as classificações (decisões) realizadas pelos classificadores individuais, são propostos os seguintes 3 (três) métodos para realizar a combinação:

1. **Unweighted Voting – UV**: \mathbf{x} é rotulado com a classe que receber mais votos dos L classificadores;
2. **Weighted by Mean Voting – WMV**: o número de votos na classe dada por cada classificador \mathbf{h}_l ao classificar \mathbf{x} é ponderado pela média estimada de erro do classificador, representado por $m_err(\mathbf{h}_l)$, e \mathbf{x} é então rotulado com a classe que tiver o peso máximo dos L classificadores:

$$WMV(\mathbf{x}, C_v) = \max_{C_i \in \{C_1, \dots, C_{N_{Cl}}\}} \sum_{l=1}^L g(\mathbf{h}_l(\mathbf{x}), C_i) , \text{ onde}$$

$$g(\mathbf{h}_l(\mathbf{x}), C_i) = \begin{cases} \lg((1 - m_err(\mathbf{h}_l))/m_err(\mathbf{h}_l)) & \text{se } \mathbf{h}_l(\mathbf{x}) = C_i, \\ 0 & \text{caso contrário.} \end{cases}$$

3. **Weighted by Mean and Standard Error Voting – WMSV**: similar ao método anterior, mas considerando também o erro padrão da média do erro do classificador, representado por $se_err(\mathbf{h}_l)$, para calcular o peso do classificador correspondente:

$$WMSV(\mathbf{x}, C_v) = \max_{C_i \in \{C_1, \dots, C_{N_{Cl}}\}} \sum_{l=1}^L g(\mathbf{h}_l(\mathbf{x}), C_i), \text{ onde}$$

$$g(\mathbf{h}_l(\mathbf{x}), C_i) = \begin{cases} \lg((1 - m_err(\mathbf{h}_l))/m_err(\mathbf{h}_l)) \\ \quad + \lg((1 - se_err(\mathbf{h}_l))/se_err(\mathbf{h}_l)) & \text{se } \mathbf{h}_l(\mathbf{x}) = C_i, \\ 0 & \text{caso contrário.} \end{cases}$$

Dado um novo exemplo a ser classificado, o *ensemble* fornece a sua classificação e o conjunto de regras que cobrem esse exemplo. Entretanto, em muitos casos esse conjunto de regras contém regras que são especializações de outras regras. Para facilitar a explicação ao usuário, decidimos fornecer como explicação ao usuário somente as regras mais gerais, por ser essa a explicação sintaticamente mais simples. Assim, propomos um método para resumir (simplificar) esse conjunto de regras, baseado nos conceitos de generalização e especialização de regras.

Os métodos para construção de *ensembles* de classificadores foram implementados, avaliados e publicados nos trabalhos [1,5,6]. Já o algoritmo para resumir a explicação do *ensemble* foi publicado em [2].

5 Algoritmos Genéticos e AM Simbólico

O principal objetivo de aprendizado de máquina é desenvolver e implementar algoritmos capazes de encontrar novos conhecimentos ou melhorar algum conhecimento previamente construído, usando observações do mundo. O interesse em aplicar algoritmos evolutivos em aprendizado de máquina tem crescido nos últimos anos. Nesses algoritmos, o conhecimento é representado por indivíduos que constituem a população do AG e podem ser manipulados por operadores genéticos.

Em aprendizado de máquina simbólico, as seguintes abordagens foram propostas pela comunidade de AGs para manipulação de conhecimento simbólico [9]: *Pittsburgh* — cada indivíduo codifica um conjunto de regras de conhecimento; e *Michigan* — cada indivíduo codifica somente uma regra. Como o objetivo da tese de doutorado foi propor um algoritmo para evoluir classificadores simbólicos, foi utilizada a abordagem Pittsburgh para evoluir classificadores simbólicos, no qual a interação entre as regras que constituem o classificador é mais importante que cada regra isolada. Exemplos de propostas de AGs que utilizam a abordagem *Pittsburgh* são [11] e [10]. O AG proposto, denominado GAESC (*Genetic Algorithm for Evolving Symbolic Classifiers*), segue os passos descritos no Algoritmo 1. A vantagem do GAESC, publicado em [3,4], em relação a outros

AGs propostos na literatura está relacionado com a codificação dos indivíduos e com a população inicial do AG. A seguir, são descritos os componentes do algoritmo GAESC.

Algoritmo 1 GAESC algorithm.

Parâmetros: $S = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_N, y_N)\}$: Conjunto de exemplos utilizados para construir os conjuntos de treinamento, teste e validação;
 $C_{init} = \{\mathbf{h}_a, \mathbf{h}_b, \dots, \mathbf{h}_q\}$;
 $\mathcal{W} = \{R_1, \dots, R_W\}$: Base de regras de conhecimento;
 N_{ind} : Número de indivíduos na população;
Parâmetros de *cross-over*, *selection* (seleção) e *mutation* (mutação);
 $f(\mathbf{h})$: Função de avaliação para evoluir os classificadores.

- 1: Gerar população inicial $P_{current} = \{\mathbf{h}_1, \dots, \mathbf{h}_{N_{ind}}\}$ onde $C_{init} \subset P_{current}$;
- 2: **for all** $\mathbf{h} \in P_{current}$ **do**
- 3: Calcular $f(\mathbf{h})$ usando o conjunto de treinamento;
- 4: **end for**
- 5: **while** Critério de parada não satisfeito **do**
- 6: $\mathbf{h} = \text{melhor_indivíduo}(P_{current})$;
- 7: $P_{new} = \mathbf{h} + N_{ind} - 1$ indivíduos selecionados de $P_{current}$;
- 8: Aplicar cross-over a pares de indivíduos selecionados de P_{new} ;
- 9: Aplicar mutação a indivíduos selecionados de P_{new} ;
- 10: Avaliar f dos indivíduos modificados de P_{new} usando o conjunto de treinamento;
- 11: $P_{current} = P_{new}$;
- 12: **end while**
- 13: $\mathbf{h} = \text{melhor_indivíduo}(P_{current})$;
- 14: $\mathbf{h}^* = \text{pós_processamento}(\mathbf{h})$; {Podar \mathbf{h} usando o conjunto de avaliação}
- 15: **return** \mathbf{h}^* ; {Classificador evoluído pelo AG}

População inicial, inicialização do AG e codificação dos indivíduos: No AG proposto, para construir os classificadores (indivíduos) iniciais, são selecionadas regras (genes) de classificação de uma base de regras. Assim, cada gene de um indivíduo (classificador) é uma regra com um identificador único, e cada indivíduo é, portanto, uma seqüência (conjunto) de identificadores de regras. Para construir a base de regras mencionada, podem ser utilizadas regras construídas por especialistas do domínio, ou regras pertencentes a classificadores induzidos por algoritmos de aprendizado simbólico, tais como $\mathcal{CN}2$, $\mathcal{C}4.5$ e $\mathcal{C}4.5rules$. Na realidade, esses classificadores podem também ser utilizados como indivíduos iniciais do AG, e são os classificadores que compõem o conjunto C_{init} no Algoritmo 1. Tais classificadores, se utilizados, podem facilitar o processo de busca do AG pela melhor solução, já que esses classificadores provavelmente são soluções melhores que as soluções construídas aleatoriamente selecionando regras da base de regras. Todas as regras presentes na base de regras devem estar em uma sintaxe padrão de regras, denominada \mathcal{PBM} .

Operadores: O operador de seleção é responsável por selecionar indivíduos para compor a nova população. O operador implementado — roleta — seleciona os

indivíduos mais aptos da população atual. É considerado mais apto o indivíduo que tiver o melhor valor da função de avaliação. O operador de *crossover* utilizado é o *crossover* assimétrico. Em um *crossover* assimétrico, a cada indivíduo da população corrente é associado aleatoriamente um valor entre 0 e 1. Os indivíduos cujo valor associado é menor que a probabilidade de *crossover*, p_c , são selecionados para a operação de *crossover*. Para cada par de indivíduos “pais” são escolhidas aleatoriamente duas posições, uma em cada indivíduo. Os indivíduos pais são divididos em dois segmentos conforme as posições escolhidas. Por fim, são criados os indivíduos filhos por meio da composição dos segmentos dos indivíduos pai. Se forem criados indivíduos filhos com duas regras idênticas, uma é removida do indivíduo resultante. O operador de mutação seleciona um gene aleatoriamente em um indivíduo e o troca por outro gene, ou seja, outra regra, selecionada aleatoriamente da base de regras disponível. A probabilidade de um indivíduo sofrer mutação é dada por p_m .

Funções de avaliação: Para avaliar os indivíduos, deve-se avaliar o comportamento do conjunto de regras (classificador) que o indivíduo representa sobre um conjunto de exemplos de teste. As medidas de avaliação de classificadores usadas no AG proposto são $HQ_{Acc}(\mathbf{h}) = Acc(\mathbf{h})$, $HQ_{Prec}(\mathbf{h}) = Prec(\mathbf{h})$, $HQ_{F1}(\mathbf{h}) = F1(\mathbf{h})$, $HQ_{AC}(\mathbf{h}) = Acc(\mathbf{h}) \times mean(Cov(R_i)), \forall R_i \in \mathbf{h}$ and $HQ_{PC}(\mathbf{h}) = Prec(\mathbf{h}) \times mean(Cov(R_i)), \forall R_i \in \mathbf{h}$. Ainda, foram utilizados os métodos *SR* e *MR* para determinar a classificação de um novo exemplo. As medidas de regra utilizadas são $Acc(R)$, $Lacc(R)$ e $Cov(R)$, definidas na Seção 3, e um índice que subtrai o erro da precisão, definido por $PE(h) = f_{bh} - f_{b\bar{h}}$. Os métodos *SR* implementados foram SR_{Acc} , SR_{Cov} , SR_{Lap} e SR_{PE} ; e os métodos *MR* implementados foram MR_{Acc} , MR_{Lap} , MR_{Cov} e MR_{PE} . Assim, as funções de avaliação são resultantes da combinação dos métodos descritos. Dessa maneira, funções de avaliação do AG proposto são $MR_{Acc}HQ_{AC}$, $MR_{Cov}HQ_{Acc}$, $SR_{Lap}HQ_{F1}$, e assim sucessivamente.

Critério de parada: São dois os critérios de parada (*SC* — *Stop Criterion*) propostos para serem utilizados no GAESC. O primeiro critério de parada, denominado SC_{Max} , é dado pelo número máximo de gerações que devem ser executadas. O segundo critério de parada, denominado SC_{Conv} é definido como segue: dado um número de gerações N_{Gen} , se a função de avaliação do melhor indivíduo não melhorar nas últimas N_{Gen} gerações, o algoritmo pára. O número de gerações de cada um dos critérios de parada são parâmetros do AG. Por *default*, N_{Gen} para o critério de parada SC_{Max} é 10, e N_{Gen} para SC_{Conv} é 5.

Pós-processamento do indivíduo resultante: Após o AG fornecer como saída o melhor indivíduo (classificador) simbólico por ele evoluído, o sistema proposto realiza um pós-processamento desse classificador. Isso é realizado da seguinte maneira: dado o conjunto de exemplos de treinamento S_{tr} , antes de iniciar a execução do AG, são retirados 10% dos exemplos desse conjunto S_{tr} , formando assim o conjunto de validação S_{val} , tal que a proporção de exemplos de S_{val} em cada classe é semelhante à do conjunto S_{tr} , sendo $S_{tr} = S_{val} \cup S'_{tr}$. Assim, o AG é executado utilizando o conjunto S'_{tr} . O conjunto S'_{tr} é utilizado a cada iteração do AG para o cálculo da função de avaliação utilizada. Ao fi-

nal da execução, o melhor indivíduo (classificador) encontrado é pós-processado utilizando todo o conjunto de exemplos de treinamento original S_{tr} . O critério de pós-processamento utilizado consiste em retirar desse classificador todas as regras que não cobrem nenhum exemplo em S_{tr} .

6 Experimentos e Resultados

Para responder às duas questões descritas na Seção 2, realizamos experimentos divididos em duas fases. Na primeira fase, para responder à primeira questão, foram utilizados 3(três) conjuntos de dados da UCI com mais de 1000 exemplos, para conseguirmos retirar amostras viáveis para se construir um classificador e avaliar a taxa de erro desse classificador utilizando 10-fold cross-validation. Os experimentos foram realizados em 5(cinco) cenários diferentes, totalizando 180 experimentos. Em cada experimento, foi induzido um *ensemble* com o total de exemplos (para avaliar também o método de simplificação da explicação, e executado 10-fold cross-validation, o que totaliza 1980 *ensembles* construídos. Nos resultados obtidos, pudemos observar que, para dois conjuntos de dados, em todos os experimentos a taxa de erro dos *ensembles* de classificadores é menor que a taxa de erro dos classificadores que os compõem. Esse resultado tem 95% de confiança segundo o teste de hipóteses t . Em relação ao terceiro conjunto de dados, somente em um cenário de experimentos, a taxa de erro dos *ensembles* de classificadores é menor que a taxa de erro dos classificadores que os compõem, com 95% de confiança segundo o teste de hipóteses t . Portanto, consideramos que a resposta para a primeira questão é afirmativa.

Na segunda fase, foram utilizados 4(quatro) conjuntos de dados, também da UCI, porém de menor porte, para tentar responder à segunda questão. Os conjuntos de dados da segunda fase foram selecionados com menor porte (menor número de exemplos) primeiro porque o algoritmo genético realiza diversas iterações, exigindo que muito tempo fosse necessário para realização de experimentos. Ainda, deve ser observado que os resultados obtidos para conjuntos de dados de menor porte pode ser estendido para conjuntos de maior porte, já que em conjuntos de maior porte basta que sejam retiradas amostras do conjunto de dados disponível que o algoritmo de aprendizado consiga manipular. Por esses motivos, avaliar o AG levaria muito tempo, e não é um fator que influencia na avaliação realizada. Para construir os *ensembles* e construir a população inicial do AG, inicialmente o conjunto de treinamento foi utilizado para induzir 3(três) classificadores, utilizando os algoritmos $CN2$, $C4.5$ e $C4.5rules$. Foram construídos *ensembles* utilizando todas as variações dos métodos de construção de *ensembles* descritos. Utilizando o teste t de hipóteses, foi verificado que não há melhoria significativa na taxa de erro dos *ensembles* de classificadores quando comparada à taxa de erro do melhor classificador, *i.e.*, do classificador com a menor taxa de erro que compõe o *ensemble*. Já para o algoritmo genético, foram avaliados os 2(dois) critérios de parada e todas as funções de avaliação foram testadas. Os experimentos foram executados em 6(seis) cenários diferentes. O que mudou em cada cenário foi o critério de convergência, o número de gerações

no critério de convergência SC_{Conv} — 5 e 10 — e a presença ou ausência na base de regras de regras construídas com o Apriori. Foi realizado um total de 960 experimentos e 9600 classificadores evoluídos com o GAESC. Foi utilizado o teste t para realizar a comparação dos resultados obtidos nos experimentos utilizado o AG com a taxa de erro do melhor classificador na população inicial, ou seja, do classificador inicial com menor taxa de erro. Considerando os resultados com todos os conjuntos de dados e em todos os cenários, é possível observar que em todos os casos, as melhores funções de avaliação foram a HQ_{Acc} e a HQ_{F1} . Também, considerando somente os resultados obtidos com HQ_{Acc} e HQ_{F1} , é possível observar que os melhores entre eles foram os obtidos com os métodos MR_{Acc} e MR_{PE} . Devido aos resultados obtidos nesta fase, consideramos que a resposta para a segunda questão também é afirmativa.

7 Conclusões

Nesta tese de doutorado, foram propostos métodos de construção de *ensembles* de classificadores simbólicos e um algoritmo genético para combinação de classificadores simbólicos. Para avaliar os métodos propostos, foram realizados experimentos divididos em duas fases. Na primeira fase de experimentos, os resultados obtidos foram bastante interessantes e promissores. Entretanto, o fato de variar os métodos de combinação de classificadores que compõem os métodos de construção de *ensembles* não ofereceu diferença na taxa de erro obtida nos *ensembles* construídos. Dessa maneira, o método de combinação sem peso pode ser mais indicado por não requerer o cálculo da estimativa de taxa de erro dos classificadores componentes, diminuindo assim o tempo de construção de *ensembles*. Em relação aos métodos de classificação de exemplos utilizados pelos classificadores componentes dos *ensembles*, os resultados obtidos com o método *SR* (*Single Rule*) não foram tão bons quanto utilizando o método *MR* (*Multiple Rule*). Em relação ao método utilizado para simplificação sintática de explicação, os resultados obtidos com o método de explicação foram muito bons, já que a quantidade de regras para explicar novas classificações após a simplificação é reduzida para mais da metade, comparada à quantidade de regras inicialmente oferecidas para a explicação da classificação dada pelo *ensemble*.

Na segunda fase de experimentos, o nosso objetivo foi testar a qualidade do indivíduo resultante da combinação dos classificadores iniciais usando o GAESC, tanto em relação a esses classificadores iniciais quanto em relação a *ensembles* construídos com esses mesmos classificadores. Os resultados obtidos com o algoritmo genético proposto são muito bons e bastante promissores, já que foram obtidos classificadores com excelentes taxas de erro, quando comparadas com as taxas de erro dos classificadores induzidos pelos algoritmos de aprendizado de máquina simbólicos utilizados, e quando comparadas com as taxas de *ensembles* construídos pelo outro método proposto. Entretanto, ao contrário do esperado, as taxas de erro obtidas com o critério de parada de convergência foram maiores que as taxas de erro obtidas utilizando o critério de parada que faz com que o AG seja executado um número fixo de iterações. Esses resultados

nos levaram a acreditar que havia a necessidade de se investigar mais profundamente o critério de convergência utilizado, bem como provavelmente propor novos critérios de parada relacionados à convergência. Em um trabalho submetido, o qual é continuação do trabalho resumidamente descrito nesta tese de doutorado, pudemos averiguar que buscar a convergência de toda a população, avaliando a convergência da média da função da avaliação de todos os indivíduos, oferecem melhores resultados. Em trabalhos futuros, pensamos ainda em explorar combinação de outras medidas de avaliação de classificadores na função de avaliação.

References

1. F. C. Bernardini and M. C. Monard. Methods for constructing symbolic ensembles from symbolic classifiers. In *Frontiers in Artificial Intelligence and Applications*, volume 132, pages 161–168, Hyogo, Japan, 2005.
2. F. C. Bernardini and M. C. Monard. Uma proposta para a construção de ensembles simbólicos que explicam suas decisões. In *Conferência Latinoamericana de Informática — CLEI 2005*, pages 151–162, Colômbia, 2005.
3. F. C. Bernardini and M. C. Monard. Uma proposta para evoluir classificadores simbólicos utilizando um algoritmo genético. In *Conferência Latinoamericana de Informática — CLEI 2006*, pages 1–12, Chile, 2006. Publicado em CD-ROM.
4. F. C. Bernardini and M. C. Monard. Utilizando a abordagem pittsburgh para evolução de classificadores simbólicos. In *WTDIA 2006*, pages 1–10, SP, Brasil, 2006. Publicado em CD-ROM.
5. F. C. Bernardini, M. C. Monard, and R. C. Prati. Constructing ensembles of symbolic classifiers. In *International Conference on Hybrid Intelligent Systems — HIS 2005*, volume 1, pages 315–320, Rio de Janeiro, 2005. California: IEEE Computer Society.
6. F. C. Bernardini, M. C. Monard, and R. C. Prati. Constructing ensembles of symbolic classifiers. *International Journal on Hybrid Intelligent Systems*, 3(3):159–167, 2006.
7. F.C. Bernardini. Combinação de classificadores simbólicos utilizando medidas de avaliação de regras e algoritmos genéticos, 2006. Tese de Doutorado, ICMC/USP. <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-29092006-110806>.
8. T. G. Dietterich. *Ensemble learning*. In M.A. Arbib, editor, *The Handbook of Brain Theory and Neural Networks, Second Edition*. MIT Press, 2002.
9. A. A. Freitas. *Data Mining and Knowledge Discovery with Evolutionary Algorithms*. Springer Verlag, 2002.
10. X. Llorà and J. M. Garrell i Guiu. Genifer: A nearest neighbour based classifier system using ga. In *Proceedings of the Genetic and Evolutionary Computation Conference – GECCO 1999*, volume 1, page 797, Orlando, Florida, 1999. Morgan Kaufmann Publishers.
11. C. Setzkorn and R. C. Paton. Merbis - a multi-objective evolutionary rule base induction system. Technical Report ULCS-03-016, Department of Computer Science, University of Liverpool, Liverpool, U.K., 2003. <http://www.csc.liv.ac.uk/research/techreports/tr2003/ulcs-03-016.ps>.