

Uma Proposta para a Construção de *Ensembles* Simbólicos que Explicam suas Decisões

Flávia Cristina Bernardini* e Maria Carolina Monard

Laboratório de Inteligência Computacional
Instituto de Ciências Matemáticas e Computação
Universidade de São Paulo
Av. do Trabalhador Sancarlense, 400 – Caixa Postal 668
CEP 13560-970 São Carlos, SP
{fbernard,mcmonard}@icmc.usp.br

Abstract

Data Mining applications generally use learning algorithms in order to induce knowledge. To accomplish this task, these algorithms should be able to operate with massive data sets. Several techniques, such as data sampling, can be used to scale up learning algorithms to deal with large datasets. In order to induce classifiers, learning algorithms can thus be applied to smaller samples of the original dataset. The individual classifiers can then be combined into ensembles which under certain conditions can be more accurate than the individual classifiers. However, when compared to symbolic classifiers, ensembles often lack the facility to explain their decisions. This work explores a method to offer concise explanation of ensembles decisions whenever the ensembles are composed by a combination of symbolic classifiers. Different methods used to construct ensembles are also described.

Keywords: Symbolic Machine Learning, Symbolic Ensembles, Combining Classifiers.

Resumo

Em aplicações práticas de Data Mining geralmente são utilizados algoritmos de aprendizado para induzir conhecimento. Entretanto, esses algoritmos nem sempre estão preparados para trabalhar com grandes bases de dados. Uma maneira de possibilitar tais algoritmos a trabalhar com grandes bases é utilizar a abordagem de *ensembles* de classificadores associada a técnicas de amostragem de conjuntos de dados. Para induzir os classificadores, algoritmos de aprendizado podem ser aplicados em amostras do conjunto de dados original. Os classificadores induzidos podem ser, então, combinados em *ensembles*, os quais podem ser, sob certas condições, mais precisos que os classificadores individuais que o compõem. Entretanto, *ensembles* frequentemente não conseguem explicar suas decisões de classificação. Neste trabalho exploramos um método para oferecer explicações concisas das decisões de classificações de *ensembles* simbólicos e também apresentamos métodos para construir esses *ensembles*.

Palavras chave: Aprendizado de Máquina Simbólico, *Ensembles* Simbólicos, Combinação de Classificadores.

1 INTRODUÇÃO

Em Aprendizado de Máquina — AM — supervisionado, além da indução de classificadores precisos, é também importante analisar o conhecimento induzido com o objetivo de descobrir novo conhecimento. Nesse caso, há um forte apelo para usar algoritmos de AM simbólicos, nos quais o modelo induzido pode ser diretamente interpretado pelo usuário/especialista do domínio. Porém, quando são utilizadas grandes bases (conjuntos) de dados, muitas vezes os algoritmos de AM simbólicos disponíveis não estão prontos para lidar com tais bases. Ainda, mesmo quando os conjuntos de dados sejam suportados por tais algoritmos, nem sempre é

*Trabalho realizado com auxílio da FAPESP, Brasil, Proc. N° 02/06914-5.

possível induzir um classificador suficientemente preciso, ou seja, que classifique novos casos com a precisão exigida pelo especialista. Uma possível solução é utilizar a abordagem de *ensembles* de classificadores, para tornar os algoritmos de AM disponíveis capazes de lidar com grandes bases e/ou melhorar o poder preditivo da classificação. Um *ensemble* consiste de um conjunto de classificadores cujas decisões individuais são combinadas de alguma maneira para classificar novos casos [9, 11].

As principais abordagens de *ensembles* encontradas na literatura apresentam comportamento tipo caixa preta, ou seja, não têm condições de oferecer uma explicação que justifique a classificação de novos exemplos [6, 12, 2, 7]. Em algumas situações, essas explicações são obrigatórias. Por exemplo, em certos países, ao informar ao cliente de um banco se seu crédito foi aprovado ou não, existe a necessidade legal de se justificar o motivo de sua aprovação ou negação. Já os *ensembles* construídos com a abordagem de construção de *ensembles* utilizada neste trabalho fornecem, para cada exemplo, a classificação desse exemplo e uma explicação relacionada à classificação realizada. Entretanto, muitas vezes essa explicação pode ser muito extensa para ser fornecida ao usuário do *ensemble* e/ou ao especialista do domínio, pois essa explicação está relacionada com as regras disparadas pelo conjunto de classificadores que constituem o *ensemble*.

O objetivo deste trabalho é propor métodos para construir *ensembles* de classificadores simbólicos bem como propor um método para resumir a explicação fornecida pelos *ensembles* construídos, apresentando assim, ao usuário do *ensemble*, uma explicação mais concisa e compreensível, relacionada à classificação de novos casos (exemplos).

O restante deste artigo está organizado como segue: na Seção 2 são descritos outros trabalhos relacionados à construção de *ensembles*; na Seção 3 são apresentados alguns conceitos para possibilitar a compreensão deste trabalho bem como a notação utilizada; na Seção 4 são descritos os métodos propostos neste trabalho para a construção de *ensembles* simbólicos; na Seção 5 é mostrado o método utilizado para resumir a explicação fornecida ao usuário pelos *ensembles* simbólicos na classificação de novos exemplos; na Seção 6 são descritos os experimentos realizados utilizando um conjunto de dados naturais e os resultados obtidos; por fim, na Seção 7 encontram-se as conclusões deste trabalho e trabalhos futuros.

2 TRABALHOS RELACIONADOS

Ensembles de classificadores têm sido bastante estudados nos últimos anos na área de AM, com a finalidade de melhorar o poder de predição de algoritmos de aprendizado supervisionado. A tarefa de construção de *ensembles* de classificadores pode ser dividida em duas sub-tarefas [9]. A primeira consiste em induzir um conjunto de classificadores componentes do *ensemble*. A segunda tarefa consiste em decidir como combinar as decisões dos classificadores componentes para classificar novos exemplos. Alguns trabalhos propõem métodos de construção de *ensembles* — *bagging* [6], *boosting* [12], dentre outros — e variações desses métodos — por exemplo, *wagging* [2] e *arcing* [7] — que envolvem ambas sub-tarefas de construção de *ensembles*. Vários trabalhos utilizam um grande número de classificadores para compor um *ensemble* [10, 2] com o objetivo de melhorar a sua precisão. Porém, em [3], é feita uma análise que explica ser possível construir bons *ensembles* de classificadores com um pequeno número de classificadores componentes. Além disso, trabalhos mais recentes têm mostrado maior preocupação com a fase de combinação de classificadores [15].

Para realizar a combinação, existem propostas de métodos de *stacking* [19], nos quais é construído um classificador para tomar a decisão final do *ensemble* dadas as decisões individuais dos classificadores. Em ambos trabalhos [15, 19], resultados satisfatórios são apresentados e é mostrado que não é necessário utilizar métodos específicos para a construção dos classificadores que compõem o *ensemble*, como os utilizados nos trabalhos iniciais em *ensembles* de classificadores [6, 12].

Os métodos de construção de *ensembles* utilizados nesses trabalhos apresentam comportamento do tipo caixa preta, ou seja, não permitem explicar as decisões de classificação tomadas pelos *ensembles* construídos. Entretanto, em algumas situações, é obrigatório o fornecimento de explicação de classificações, como citado anteriormente. Porém, não é de nosso conhecimento trabalhos relacionados com o tratamento de explicação de decisões tomadas por *ensembles* de classificadores simbólicos, o qual é o foco deste trabalho.

3 CONCEITOS E NOTAÇÃO

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 — Tabela 1.

	X_1	X_2	...	X_M	Y
T_1	x_{11}	x_{12}	...	x_{1M}	y_1
T_2	x_{21}	x_{22}	...	x_{2M}	y_2
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
T_N	x_{N1}	x_{N2}	...	x_{NM}	y_N

Tabela 1: Conjunto de exemplos no formato atributo-valor

Em problemas de classificação, tratados neste trabalho, o atributo classe y_i é discreto, ou seja, $y_i \in \{C_1, C_2, \dots, C_{N_{C_i}}\}$. A partir do conjunto de treinamento S , um *classificador* \mathbf{h} é induzido. No caso de aprendizado simbólico proposicional, \mathbf{h} pode ser transformado em um conjunto de regras **if-then**, não ordenadas ou disjuntas, tal que $\mathbf{h} = \{R_1, R_2, \dots, R_{N_R}\}$. Neste trabalho, as regras que participam de \mathbf{h} são regras de classificação, definidas mais adiante, ainda nesta seção.

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 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.

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 . 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 $C_i \in \{C_1, \dots, C_{N_{C_i}}\}$. A maioria dos algoritmos de indução de regras de classificação pertencem a uma das famílias de algoritmos: “separar-para-conquistar” e “dividir-para-conquistar”. Algoritmos da primeira família geralmente usam um algoritmo iterativo e guloso de cobertura sobre o conjunto de exemplos de treinamento. A cada iteração, o algoritmo encontra a melhor regra e remove os exemplos cobertos por essa regra. Finalmente, um classificador é construído com as regras encontradas. Este classificador é composto por uma lista de regras ordenadas (ou lista de decisão), caso tiverem sido removidos, em cada iteração do algoritmo, todos os exemplos cobertos pela regra descoberta, ou é composto por um conjunto de regras não ordenadas, caso tiverem sido removidos, em cada iteração do algoritmo, somente os exemplos corretamente cobertos pela regra. Algoritmos da segunda família — “dividir-para-conquistar” — constroem um classificador global utilizando uma estratégia *top-down*, refinando, a cada iteração, a teoria parcial. Geralmente, classificadores induzidos por algoritmos dessa família são expressos como árvores de decisão, as quais podem ser escritas como um conjunto de regras disjuntas e não ordenadas.

Dadas duas regras de classificação $R_i = B_i \rightarrow H_i$ e $R_j = B_j \rightarrow H_j$ com cabeças iguais ($H_i = H_j$), é muito simples determinar se R_i é uma generalização de R_j , *i.e.* R_i *subsume* R_j , ou se R_i é uma especialização de R_j , *i.e.* R_j *subsume* R_i , considerando o conjunto de testes de atributos que participam do corpo dessas regras. Sejam A_i e A_j , respectivamente, os conjuntos de testes de atributos que participam nos corpos B_i e B_j das regras R_i e R_j respectivamente. Então, R_i é uma generalização de R_j , ou R_j é uma especialização de R_i se $A_i \subseteq A_j$. Por exemplo, considerando as duas regras de classificação:

$$R_i = \text{If } at_1 = 2 \text{ and } at_2 > 4 \text{ then class} = +$$

e

$$R_j = \text{If } at_1 = 2 \text{ and } at_2 > 4 \text{ and } at_3 = 5 \text{ then class} = +$$

então

$$A_i = \{at_1 = 2, at_2 > 4\}$$

e

$$A_j = \{at_1 = 2, at_2 > 4, at_3 = 5\}.$$

Assim, R_i é uma generalização de R_j pois $A_i \subseteq A_j$ ¹. Com a finalidade de simplificar uma explicação composta por um conjunto de regras \mathbf{R}_{expl} , a explicação continua sendo válida removendo-se de $\mathbf{R}_{expl} = \{R_1, \dots, R_P\}$ as regras que são especialização de outras regras também pertencentes a \mathbf{R}_{expl} porque, dadas duas regras R_i e R_j , e dado que R_i é uma generalização de R_j , os exemplos cobertos por R_j são cobertos pela regra mais geral R_i . Essa idéia é utilizada neste trabalho para simplificar a explicação de *ensembles* simbólicos, como explicado na Seção 5.

4 COMBINAÇÃO DE CLASSIFICADORES SIMBÓLICOS

Como mencionado, o processo de construção de um *ensemble* de classificadores pode ser dividido em duas sub-tarefas [11]:

1. indução de um conjunto de classificadores; e
2. combinação das decisões dos classificadores induzidos para classificar novos exemplos.

Neste trabalho, a primeira tarefa 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 $\mathbf{h}_1, \dots, \mathbf{h}_L$. Deve ser observado que o algoritmo não precisa ser o mesmo para todas as L amostras, qualquer algoritmo de aprendizado simbólico pode ser utilizado. A única restrição é quanto ao tipo de regras geradas pelo algoritmo, as quais devem ser regras não ordenadas, descritas na Seção 3. 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. Na Figura 1 é ilustrado o método proposto, onde $Combinar(\mathbf{h}_1(\mathbf{x}), \dots, \mathbf{h}_L(\mathbf{x}))$ constitui o *ensemble* simbólico.

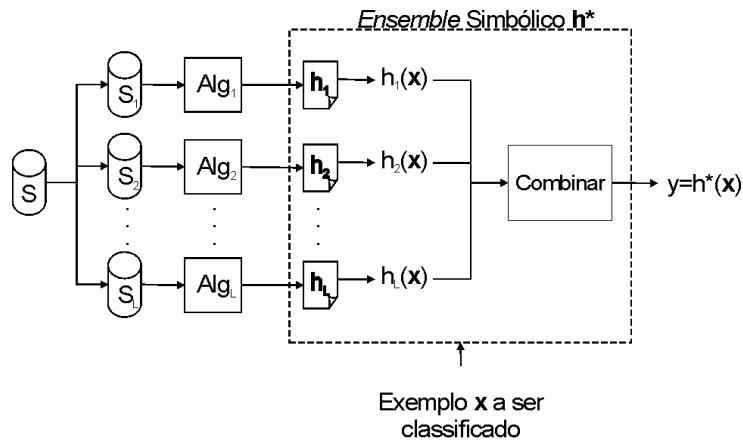


Figura 1: Um método para construção de *ensembles* de classificadores.

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;

¹Deve ser observado que $A_i \subseteq A_j$ considera que para testes de um mesmo atributo envolve análise de subconjuntos quando o atributo X_i *op Valor*, o operador “ \leq ” é uma generalização de “ $<$ ”, “ \geq ” é uma generalização de “ $>$ ”, e assim sucessivamente.

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

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

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:

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

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

Com o objetivo de avaliar essas propostas, foi implementado um sistema computacional denominado Ensemble Learning Environment (ELE), integrado ao ambiente computacional DISCOVER. O ambiente DISCOVER tem como principal objetivo integrar e padronizar os diversos projetos desenvolvidos no Laboratório de Inteligência Computacional – LABIC² – do ICMC-USP, relacionados com pré-processamento de conjuntos de dados, aquisição automática de conhecimento e avaliação de conhecimento [17]. Na maioria desses projetos, incluindo o projeto ELE, diversas tarefas tais como transformação de dados e formatos, execução de algoritmos, medições, entre outras, devem ser executadas diversas vezes. Assim, muitas ferramentas foram e estão sendo implementadas em *Perl* [20] no DISCOVER, como bibliotecas de classes, para automatizar parcial ou integralmente algumas dessas tarefas. O ambiente DISCOVER oferece vantagens em relação a outros sistemas com objetivos semelhantes, pois permite a visão unificada que os formatos baseados em padrões proporcionam ao pesquisador (desenvolvedor) de novos componentes. Os padrões de representação foram sendo definidos por área, sendo que em [16] é proposta uma sintaxe padrão para representação de conhecimento de diversos indutores simbólicos denominada *PBM*. Para a representação de dados foi proposta uma sintaxe padrão, denominada DSX – *Discover Dataset Syntax*, a qual permite a utilização da biblioteca de classes DOL [1], para converter os arquivos de dados para sintaxe utilizada por diversos sistemas de aprendizado simbólico, tais como *C4.5*, *C4.5rules*, *CN2* entre outros. O sistema ELE, no qual estão implementados os métodos de combinação de classificadores UV, WMV e WMSV, foi desenvolvido como uma biblioteca de classes do DISCOVER utilizando várias de suas funcionalidades.

Em [4], são descritos experimentos realizados com o sistema ELE utilizando o conjunto de dados Nursery, da UCI [13]. Os resultados obtidos foram satisfatórios, considerando que as estimativas de erro dos *ensembles* de classificadores foram menores que as estimativas de erro dos classificadores que o compõem. Entretanto, como mencionado anteriormente, além da preocupação com o poder de predição em problemas de AM em termos de precisão de classificação, existe também a preocupação em fornecer ao usuário a explicação da predição feita pelo sistema. Neste trabalho, são apresentados os resultados obtidos nos experimentos realizados com o conjunto de dados Nursery, tanto em relação às estimativas de erro dos *ensembles* — os quais encontram-se melhor detalhados em [4] — quanto em relação à explicação relacionada com a classificação dada pelo *ensemble*, foco deste trabalho. Na próxima seção é descrito um algoritmo que resume as regras fornecidas pelo *ensemble* utilizado para predizer um dado exemplo, de maneira a mostrar ao usuário uma explicação mais concisa e, portanto, mais fácil de ser compreendida pelo usuário do sistema.

5 CONSTRUÇÃO DA EXPLICAÇÃO

Dado um *ensemble* de classificadores simbólicos e um exemplo \mathbf{x} a ser classificado, o sistema ELE pode utilizar uma das 3 (três) maneiras descritas na seção anterior para combinar as decisões de cada um dos

²<http://labic.icmc.usp.br>

classificadores componentes. Como os classificadores são simbólicos, o *ensemble* oferece como saída, além da classificação final do exemplo \mathbf{x} , todas as regras que cobrem esse exemplo. Dessas regras, são selecionadas aquelas cuja cabeça (classe) H_i prediz para o exemplo \mathbf{x} a mesma classe que o *ensemble*, as quais formam o conjunto de regras \mathbf{R}_{expl} . Entretanto, nesse conjunto podem existir regras mais gerais que outras e, como mencionado na Seção 3, para tornar a explicação mais simples, retira-se de \mathbf{R}_{expl} as regras mais especializadas em relação a outras também pertencentes a \mathbf{R}_{expl} . Ou seja, analisando as regras em \mathbf{R}_{expl} duas a duas, ficam no conjunto final \mathbf{R}'_{expl} somente as regras que subsumem outras regras.

O procedimento para encontrar a explicação final \mathbf{R}'_{expl} é o seguinte: o corpo de cada regra R_e pertencente a \mathbf{R}_{expl} , B_e , consiste de uma conjunção de testes de atributos; cada corpo, então, é transformado no conjunto correspondente A_e de testes de atributos. O conjunto composto por esses conjuntos A_e é denominado \mathbf{A}_{expl} . Após construído o conjunto \mathbf{A}_{expl} , o Algoritmo 1 descreve como é construída a explicação mais geral da classificação atribuída pelo *ensemble* ao exemplo \mathbf{x} , a qual será mostrada ao usuário.

Algoritmo 1 Simplificação de Explicação

Require: \mathbf{x} : exemplo classificado pelo *ensemble*;

$\mathbf{R}_{expl} = \{\mathbf{R}_i, \dots, \mathbf{R}_j, \dots, \mathbf{R}_q\}$: conjunto de regras dos classificadores $\mathbf{h}_1, \dots, \mathbf{h}_L$ que constituem o *ensemble*, cobrem \mathbf{x} e predizem a classe $\mathbf{h}^*(\mathbf{x})$.

1: **procedure** explicacao($\mathbf{x}, \mathbf{R}_{expl}, \mathbf{A}_{expl}$)

2: $\mathbf{A}_{expl} = \{A_i, \dots, A_j, \dots, A_q\}$; {conjunto de conjuntos de testes de atributos A_e correspondente às regras R_e do *ensemble* que classificam o exemplo como a mesma classe $\mathbf{h}^*(\mathbf{x})$ do *ensemble*.}

3: **if** $|\mathbf{A}_{expl}| > 1$ **then**

4: **for all** $(A_i, A_j) | i \neq j$ **do**

5: **if** $A_i \subseteq A_j$ **then**

6: $\mathbf{A}_{expl} = \mathbf{A}_{expl} - \{A_j\}$

7: **end if**

8: **end for**

9: **end if**

10: **return** \mathbf{A}_{expl} ;

6 EXPERIMENTOS E RESULTADOS

Conforme mencionado anteriormente, em [4] são descritos experimentos realizados com o sistema ELE utilizando o conjunto de dados Nursery, da UCI [13]. Esse conjunto de dados foi extraído de um modelo de decisão hierárquica originalmente desenvolvido para classificar pedidos em berçários; foi utilizado durante a década de 80 quando existiam muitas inscrições nesses tipos de escolas em Ljubljana, Slovenia, e as inscrições negadas necessitavam de uma explicação objetiva. A decisão final depende de 3 sub-problemas: ocupação dos pais, estrutura familiar, posição social, e quadro de saúde e social da família da criança inscrita.

Na Tabela 2 são mostradas algumas característica do conjunto de dados Nursery: número de exemplos (# Ex.); número de atributos (contínuos, discretos) (# Atr.); distribuição de exemplos nas classes (Classe %) e erro majoritário do conjunto de dados. Deve ser observado que tal conjunto não possui valores desconhecidos nem exemplos duplicados ou conflitantes.

# Ex.	# Atr. (cont.,disc.)	Classe	Classe %	Erro Majoritário
12960	8 (0,8)	<i>not_recom</i>	33.33%	66.67% em <i>not_recom</i>
		<i>recommend</i>	0.02%	
		<i>very_recom</i>	2.53%	
		<i>priority</i>	32.92%	
		<i>spec_prior</i>	31.20%	

Tabela 2: Sumário das características do conjunto de dados Nursery

Na Tabela 3, são descritos os atributos do conjunto de dados.

Diversos experimentos foram realizados utilizando os algoritmos de aprendizado simbólico CN^2 [8] e $C4.5$ [18] para induzir os classificadores que compõem os *ensembles* [4]. A seguir são mostrados os resultados

Atributo	Descrição
parents	Ocupação dos pais (usual, pretentious, great_pret).
has_nurs	Berçário da criança (proper, less_proper, improper, critical, very_crit).
form	Estrutura da família da criança (complete, completed, incomplete, foster).
children	Número de crianças (1, 2, 3, more).
housing	Condição residencial (convenient, less_conv, critical).
finance	Condição financeira da família (convenient, inconv).
social	Condições sociais da família (non_prob, slightly_prob, problematic).
health	Condições de saúde da família (recommended, priority, not_recom).

Tabela 3: Descrição dos atributos do conjunto de dados Nursery

de 5 (cinco) desses experimentos, os quais foram realizados variando-se o número de amostras de exemplos (retiradas do conjunto de exemplos inicial sem reposição), ou seja, variando-se o número de classificadores que compõem os *ensembles* e variando-se o algoritmo de aprendizado usado em cada amostra para induzir o classificador componente, como é mostrado na Tabela 4. Por exemplo, no primeiro experimento (*Exp 1*), 3 (três) amostras foram retiradas do conjunto inicial e foi utilizado o algoritmo $\mathcal{CN}2$ para induzir os 3 classificadores componentes, enquanto que no experimento *Exp 4*, foram retiradas 5 (cinco) amostras, sendo que $\mathcal{CN}2$ foi utilizado em 3 (três) amostras e $\mathcal{C4.5}$ em 2 (duas) amostras³.

Experimento	# de Classificadores	Algoritmos de AM
<i>Exp 1</i>	3	$\mathcal{CN}2$ - $\mathcal{CN}2$ - $\mathcal{CN}2$
<i>Exp 2</i>	3	$\mathcal{C4.5}$ - $\mathcal{C4.5}$ - $\mathcal{C4.5}$
<i>Exp 3</i>	5	$\mathcal{CN}2$ - $\mathcal{CN}2$ - $\mathcal{CN}2$ - $\mathcal{CN}2$ - $\mathcal{CN}2$
<i>Exp 4</i>	5	$\mathcal{CN}2$ - $\mathcal{CN}2$ - $\mathcal{CN}2$ - $\mathcal{C4.5}$ - $\mathcal{C4.5}$
<i>Exp 5</i>	5	$\mathcal{C4.5}$ - $\mathcal{C4.5}$ - $\mathcal{C4.5}$ - $\mathcal{C4.5}$ - $\mathcal{C4.5}$

Tabela 4: Descrição dos experimentos

Na Tabela 5 são mostrados os resultados obtidos nesses 5 (cinco) experimentos. Inicialmente, para cada amostra S_i utilizada no experimento, é mostrada a taxa de erro do classificador induzido com essa amostra, e o erro padrão entre parênteses. Tal taxa de erro foi estimada com a técnica de *10-fold cross-validation* estratificado. Na seqüência, são mostrados os resultados obtidos com cada um dos três métodos de combinação de classificadores utilizados, ou seja, a taxa de erro do *ensemble* final, também estimado com a técnica de *10-fold cross-validation* estratificado. Pode ser observado na Tabela 5 que 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*. Também é interessante observar que o erro dos *ensembles* é menor, ainda quando usado o mesmo algoritmo de aprendizado (*Exp 1*, *Exp 2*, *Exp 3* e *Exp 5*), ou seja, quando a variabilidade dos classificadores induzidos é mais restrita.

Com relação aos três métodos de combinação utilizados, *i.e* UV, WMV e WMSV, não foi observada diferença significativa entre eles para este conjunto de dados. Ainda, nos experimentos *Exp 2* e *Exp 5*, nos quais todos os classificadores foram induzidos com $\mathcal{C4.5}$, foram observados resultados idênticos entre os três métodos de combinação propostos. Em todos os experimentos realizados, o algoritmo $\mathcal{C4.5}$ foi utilizado com a opção de induzir árvores de decisão, as quais foram transformadas em regras **if-then**. Os resultados na Tabela 5 mostram que, utilizando somente árvores de decisão para compor o *ensemble*, variar o método de combinação de classificadores dentre os três propostos não altera a taxa de erro dos *ensembles* construídos para este conjunto de dados, pois, como mencionado na Seção 3, as regras que compõem uma árvore de decisão são disjuntas.

Para avaliar o método de simplificação de explicação proposto neste trabalho, foram realizados experimentos com o método de simplificação de explicação nos *ensembles* construídos com os 3 (três) métodos de combinação propostos — UV, WMV e WMSV — para os 5 (cinco) experimentos — Tabela 4.

³O conjunto de dados utilizado para ilustrar o sistema é de médio porte. Assim, ele pode ser manipulado diretamente por ambos os algoritmos de aprendizado. A título de informação, a taxa de erro (e seu respectivo erro padrão) utilizando $\mathcal{CN}2$ e $\mathcal{C4.5}$ em todo o conjunto de dados é, respectivamente, 2.02 (0.19) e 2.92 (0.19)

Amostra	<i>Exp 1</i>	<i>Exp 2</i>	<i>Exp 3</i>	<i>Exp 4</i>	<i>Exp 5</i>
S_1	5.86 (0.34)	6.25 (0.26)	7.79 (0.17)	7.90 (0.15)	7.81 (0.25)
S_2	5.70 (0.26)	6.03 (0.22)	8.01 (0.19)	7.82 (0.27)	7.63 (0.26)
S_3	6.06 (0.17)	6.27 (0.14)	7.74 (0.19)	7.54 (0.21)	7.66 (0.25)
S_4	- -	- -	8.19 (0.15)	7.82 (0.30)	7.75 (0.23)
S_5	- -	- -	7.90 (0.27)	7.83 (0.35)	8.07 (0.28)
UV	4.38 (0.15)	4.95 (0.17)	5.68 (0.16)	5.31 (0.19)	6.39 (0.14)
WMV	4.31 (0.15)	4.95 (0.17)	5.45 (0.19)	5.19 (0.14)	6.39 (0.14)
WMSV	4.17 (0.21)	4.95 (0.17)	5.42 (0.21)	5.33 (0.17)	6.39 (0.14)

Tabela 5: Resultados obtidos com o conjunto de dados Nursery na fase de construção dos *ensembles* de classificadores.

Para avaliar a performance do algoritmo de simplificação de explicação, foi analisada a taxa de redução do número de regras entre o conjunto \mathbf{A}_{expl} inicial e o conjunto \mathbf{A}'_{expl} final construído utilizando o Algoritmo 1. Esse conjunto final será denominado \mathbf{A}'_{expl} . Dado um exemplo \mathbf{x} , a taxa de redução no número de regras entre o conjunto inicial $\mathbf{A}_{expl}(\mathbf{x})$ e o conjunto final $\mathbf{A}'_{expl}(\mathbf{x})$ da explicação da classificação do exemplo \mathbf{x} , $TR(\mathbf{x})$, é definida pela Equação 1.

$$TR(\mathbf{x}) = \frac{|\mathbf{A}_{expl}(\mathbf{x})| - |\mathbf{A}'_{expl}(\mathbf{x})|}{|\mathbf{A}_{expl}(\mathbf{x})|}. \quad (1)$$

Para obter a média e o erro padrão da taxa de redução para cada *ensemble* construído, foram utilizadas as taxas de redução de cada exemplo no conjunto de dados Nursery. Na Tabela 6 são mostrados os resultados obtidos nessa fase de simplificação. Especificamente, são mostrados o número médio de regras em \mathbf{A}_{expl} e \mathbf{A}'_{expl} e o erro padrão correspondente, bem como a média da taxa de redução nesse conjunto de dados.

Como pode ser observado na Tabela 6, os resultados obtidos são bastante significativos, tendo em vista que há uma redução no número de regras de, na média, mais de 50% em todos os experimentos realizados, ou seja, o número de regras de explicação para cada exemplo foi, na média, reduzido pela metade, sendo que em muitos casos, é observada uma redução muito maior. Verificando as regras que cobrem cada exemplo \mathbf{x} , *i.e.* $\mathbf{A}_{expl}(\mathbf{x})$ e $\mathbf{A}'_{expl}(\mathbf{x})$, foi observado que realmente foram removidas regras mais específicas e não somente regras idênticas⁴. Em relação às classes com pequena distribuição de exemplo, para a classe *very-recom* (2.53% dos exemplos — Tabela 2), os exemplos dessa classe foram corretamente classificados pelo *ensemble*. Já em relação à classe *recommend* (0.02% dos exemplos — Tabela 2), como o número de exemplos é excessivamente pequeno, devido ao método de amostragem utilizado (sem reposição), o *ensemble* não conseguiu classificar exemplos dessa classe.

Pode ser também observado na Tabela 6 que nos experimentos *Exp 2* e *Exp 5*, nos quais todos os classificadores que constituem o *ensemble* são induzidos utilizando $\mathcal{C}4.5$, os resultados para os três métodos de combinação são iguais. Esse resultado era esperado já que existe no máximo somente uma regra que cobre o exemplo, pois foi utilizada a opção de indução de árvores de decisão com esse algoritmo de aprendizado ($\mathcal{C}4.5$). Isso também pode ser observado nos experimentos realizados na fase de construção dos *ensembles* (*Exp 2* e *Exp 5* — Tabela 5).

Com o objetivo de ilustrar a simplificação de explicação, nas Tabelas 7 e 8 são mostrados os resultados de simplificação de explicação para 2 (dois) exemplos diferentes no experimento *Exp 1* utilizando o método de combinação UV. Nessas tabelas, na primeira linha é mostrado o exemplo a ser classificado e a classe verdadeira do exemplo. Na segunda linha encontra-se a classificação fornecida pelo *ensemble*. Após, para

⁴Na realidade, na implementação do Algoritmo 1, \mathbf{A}_{expl} é construído utilizando os corpos de todas as regras correspondentes, ou seja, das regras pertencentes ao *ensemble* que cobrem o exemplo \mathbf{x} cuja cabeça prediz a classe $\mathbf{h}^*(\mathbf{x})$.

Método de Combinação		<i>Exp 1</i>	<i>Exp 2</i>	<i>Exp 3</i>	<i>Exp 4</i>	<i>Exp 5</i>
\mathbf{A}_{expl}	UV	4.04 (2.20)	2.92 (0.26)	6.52 (3.22)	5,77 (2,06)	4,78 (0,55)
	WMV	4.04 (2.20)	2.92 (0.26)	6.51 (3.24)	5,77 (2,08)	4,78 (0,55)
	WMSV	4.04 (2.20)	2.92 (0.26)	6.51 (3.25)	5,77 (2,07)	4,78 (0,55)
\mathbf{A}'_{expl}	UV	1.70 (1.11)	1.17 (0.38)	2.01 (1.41)	1,80 (1,16)	1,16 (0,40)
	WMV	1.69 (1.12)	1.17 (0.38)	2.00 (1.42)	1,80 (1,16)	1,16 (0,40)
	WMSV	1.69 (1.12)	1.17 (0.38)	2.00 (1.42)	1,80 (1,16)	1,16 (0,40)
$TR\%$	UV	58.19 (16.19)	59.43 (14.22)	68.88 (17.46)	68,88 (16,25)	75,02 (10,61)
	WMV	58.18 (16.22)	59.43 (14.22)	68.76 (17.75)	68,79 (16,49)	75,02 (10,61)
	WMSV	58.18 (16.21)	59.43 (14.22)	68.77 (17.39)	68,92 (16,15)	75,02 (10,61)

Tabela 6: Resultados obtidos com o conjunto de dados Nursery na fase de simplificação de explicação.

cada um dos classificadores que compõem o *ensemble*, são mostrados os corpos das regras que cobrem o exemplo com a classe determinada pelo *ensemble*. Finalmente, é mostrado o conjunto dos corpos das regras a ser mostrado ao usuário para explicar a classificação do *ensemble* para esse exemplo. Deve ser observado que corpos de regras iguais são enumerados com o mesmo valor numérico (índice) e, caso uma regra for especialização de outra regra, o corpo da regra é enumerado com o índice da regra mais geral seguido do símbolo *. Por exemplo, na Tabela 7, os dois corpos de regras do classificador 2 e o corpo da regra do classificador 3 identificados por 4* são especializações do corpo da regra 4.

Nas Tabelas 7 e 8, pode-se observar que regras especializadas de outras são induzidas, e que o algoritmo para simplificação de explicação é bastante útil para usuários do sistema ELE, já que diminui bastante a quantidade de regras que explicam ao usuário/especialista a decisão tomada pelo *ensemble*.

É interessante notar nos resultados descritos nessas tabelas que, ainda que foi utilizado o mesmo algoritmo de AM ($\mathcal{CN}2$) no experimento *Exp 1*, de onde foram extraídos esses resultados, são induzidas algumas regras distintas nos diferentes classificadores. Isso se deve ao fato do $\mathcal{CN}2$ induzir regras não ordenadas e não disjuntas, ao contrário das regras que representam a árvore de decisão induzidas pelo $\mathcal{C4.5}$, que também são não ordenadas mas são disjuntas.

7 CONCLUSÕES E TRABALHOS FUTUROS

Neste trabalho foram propostos três métodos de construção de *ensembles* de classificadores simbólicos e um método para simplificar a explicação fornecida pelos *ensembles* simbólicos construídos com os métodos aqui propostos. Também, foram descritos os sistemas computacionais implementados para avaliar esses métodos.

Foi conduzida uma série de experimentos usando um conjunto de dados da UCI com tamanho limitado. Apesar de para este conjunto de dados não haver diferença significativa entre os três métodos de combinação propostos, a taxa de erro estimada dos *ensembles* construídos é menor que a taxa de erro de cada um dos classificadores que o compõem, com 95% de confiança segundo o teste de hipóteses *t*. Assim, é interessante observar que a técnica, a qual contempla o uso de grandes bases de dados, também pode ser utilizada para extrair conhecimento de conjuntos de dados de tamanho menor. Ainda, em todos os experimentos realizados com o método de simplificação de explicação, houve uma taxa significativa de redução de regras. Esses resultados são muito encorajadores.

Como trabalhos futuros, serão conduzidos mais experimentos utilizando outros conjuntos de dados, tanto de grande quanto de pequeno porte, variando o método de amostragem, o número L de classificadores induzidos e os algoritmos de aprendizado simbólicos utilizados para induzir os classificadores do *ensemble*.

Exemplo x_1 : (usual,proper,complete,1,convenient,convenient,problematic,priority) - Classe: priority

Método de Combinação: UV - Experimento: *Exp 1* - Classificação: priority

Corpo das regras com classe igual a priority que cobrem x_1 (que compõem A_{expl}):

Classificador 1:

- 1 parents = usual AND has_nurs = proper AND health = priority
- 2 has_nurs = proper AND housing = convenient AND finance = convenient AND health = priority
- 3 form = complete AND children = 1 AND housing = convenient AND finance = convenient AND health = priority
- 4 has_nurs = proper AND children = 1 AND health = priority

Classificador 2:

- 1 parents = usual AND has_nurs = proper AND health = priority
- 3 form = complete AND children = 1 AND housing = convenient AND finance = convenient AND health = priority
- 2 has_nurs = proper AND housing = convenient AND finance = convenient AND health = priority
- 4* has_nurs = proper AND form = complete AND children = 1 AND health = priority
- 4* has_nurs = proper AND children = 1 AND housing = convenient AND social = problematic AND health = priority

Classificador 3:

- 1 parents = usual AND has_nurs = proper AND health = priority
- 2 has_nurs = proper AND housing = convenient AND finance = convenient AND health = priority
- 4* has_nurs = proper AND form = complete AND children = 1 AND health = priority
- 3 form = complete AND children = 1 AND housing = convenient AND finance = convenient AND health = priority

Corpo das regras para explicação ao usuário (que compõem A'_{expl}):

- 1 parents = usual AND has_nurs = proper AND health = priority
- 2 has_nurs = proper AND housing = convenient AND finance = convenient AND health = priority
- 3 form = complete AND children = 1 AND housing = convenient AND finance = convenient AND health = priority
- 4 has_nurs = proper AND children = 1 AND health = priority

Tabela 7: Resultados obtidos com o conjunto de dados Nursery na fase de simplificação de explicação — Exemplo x_1 .

Exemplo x_2 : (usual,proper,complete,1,less_conv,inconv,problematic,priority) - Classe: priority

Método de Combinação: UV - Experimento: *Exp 1* - Classificação: priority

Corpo das regras com classe igual a priority que cobrem x_2 (que compõem A_{expl}):

Classificador 1:

- 1 parents = usual AND has_nurs = proper AND health = priority
- 2 has_nurs = proper AND children = 1 AND health = priority

Classificador 2:

- 1 parents = usual AND has_nurs = proper AND health = priority
- 2* has_nurs = proper AND children = 1 AND housing = less_conv AND health = priority
- 2* has_nurs = proper AND form = complete AND children = 1 AND health = priority

Classificador 3:

- 1 parents = usual AND has_nurs = proper AND health = priority
- 2* has_nurs = proper AND form = complete AND children = 1 AND health = priority

Corpo das regras para explicação ao usuário (que compõem A'_{expl}):

- 1 parents = usual AND has_nurs = proper AND health = priority
- 2 has_nurs = proper AND children = 1 AND health = priority

Tabela 8: Resultados obtidos com o conjunto de dados Nursery na fase de simplificação de explicação — Exemplo x_2 .

Com relação à explicação, atualmente o sistema mostra ao usuário/especialista o conjunto de regras diferentes e mais gerais que cobrem o exemplo com a classe determinada pelo *ensemble*. Entretanto, após encontrado esse conjunto de regras, é possível ordenar essas regras utilizando diferentes medidas de avaliação de regras, tais como precisão, sensibilidade, especificidade, suporte, novidade e outras [14, 5]. Dessa forma, o usuário poderia especificar uma medida de avaliação a ser utilizada e essas regras seriam mostradas em ordem de importância segundo essa medida. Essa extensão será implementada futuramente.

AGRADECIMENTOS

Os autores agradecem as valiosas sugestões dos revisores.

Referências

- [1] G. E. A. P. A. Batista and M. C. Monard. Descrição da arquitetura e do projeto do ambiente computacional DISCOVER Learning Environment - DLE. Technical Report 187, ICMC/USP, 2003. ftp://ftp.icmc.usp.br/pub/BIBLIOTECA/rel_tec/RT_187.PDF.
- [2] E. Bauer and R. Kohavi. An empirical comparison of voting classification algorithms: Bagging, boosting and variants. *Machine Learning*, 36(1/2):105–139, 1999.
- [3] F. C. Bernardini. Combinação de classificadores utilizando medidas de avaliação de regras e algoritmos evolutivos, 2003. Exame de Qualificação de Doutorado, ICMC/USP.
- [4] F. C. Bernardini and M. C. Monard. Methods for constructing symbolic ensembles from symbolic classifiers. In *Fifth Congress of Logic Applied to Technology – LAPTEC 2005 – Advances in Intelligent Systems and Robotics*, Japão, 2005. (in print).
- [5] F. C. Bernardini, M. C. Monard, H. D. Lee, and S. Esteves. Um algoritmo para selecionar regras de conhecimento utilizando medidas de avaliação de regras. In *IV Workshop de Inteligência Artificial – Jornadas Chilenas de Computacion*, pages 1–7, Chillan, Chile, 2003.
- [6] L. Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.
- [7] L. Breiman. Arcing classifiers. *The Annals Of Statistics*, 26(3):801–849, 1998.
- [8] P. Clark and T. Niblett. The CN^2 induction algorithm. *Machine Learning*, 3(4):261–283, 1989.
- [9] T. G. Dietterich. Machine learning research: Four current directions. *AI Magazine*, 18:97–136, 1997. <http://www.cs.orst.edu/~tgd/>.
- [10] T. G. Dietterich. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting and randomization. *Machine Learning*, 40:139–157, 2000. <http://www.cs.orst.edu/~tgd/>.
- [11] T. G. Dietterich. Ensemble methods in machine learning. In *First International Workshop on Multiple Classifier Systems. Lecture Notes in Computer Science*, volume 1857, pages 1–15, New York, 2000. Springer Verlag.
- [12] Y. Freund and R.E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.
- [13] S. Hettich, C. L. Blake, and C. J. Merz. UCI repository of machine learning databases, 1998. <http://www.ics.uci.edu/~mllearn/MLRepository.html>.
- [14] N. Lavrac, P. Flach, and B. Zupan. Rule evaluation measures: a unifying view. In *Proc. 9th International Workshop on Inductive Logic Programming. Lecture Notes in Artificial Intelligence*, volume 1634, pages 74–185, New York, 1999. Springer Verlag.
- [15] K. T. Leung and D. S. Parker. Empirical comparisons of various voting methods in bagging. In *KDD '03: Proc. 9th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 595–600. ACM Press, 2003.

- [16] R. C. Prati, J. A. Baranauskas, and M. C. Monard. Extração de informações padronizadas para a avaliação de regras induzidas por algoritmos de aprendizado de máquina simbólico. Technical Report 145, ICMC/USP, 2001. ftp://ftp.icmc.sc.usp.br/pub/BIBLIOTECA/rel_tec/RT_145.ps.zip.
- [17] R. C. Prati, M. R. Geromini, and M. C. Monard. A framework to integrate data mining components. In *XXIX Conferencia Latinoamericana de Informatica — CLEI2003*, pages 1–11, Bolivia, 2003.
- [18] J. R. Quinlan. *C4.5 Programs for Machine Learning*. Morgan Kaufmann Publishers, Inc., 1988.
- [19] L. Todorovski and S. Dzeroski. Combining classifiers with meta decision trees. *Machine Learning*, 50(3):223–249, 2003.
- [20] L. Wall, T. Christiansen, and R. Schwartz. *Programming in PERL*. Morgan Kaufmann Publishers, Inc., 1999.