UM ALGORITMO GENÉTICO PARA DESCOBRIR REGRAS DE CLASSIFICAÇÃO EM *DATA MINING*

MARCOS VINICIUS FIDELIS[1], HEITOR SILVÉRIO LOPES[2], ALEX ALVES FREITAS[3]

[1]UEPG, CPD, Praça Santos Andrade, s/n - Ponta Grossa, PR, Brasil - 80010-330
[1]CEFET-PR, Unidade de PG, Av. Monteiro Lobato Km 4, s/n - Ponta Grossa, PR, Brasil - 84000-000
[2]CEFET-PR, CPGEI, Av. Sete de Setembro, 3165 - Curitiba, PR, Brasil - 80230-901
[3]PUC-PR, PPGIA, Rua Imaculada Conceição, 1155, Prado Velho - Curitiba, PR, Brasil - 80215-901
[1]fidelis@cpgei.cefetpr.br, [2]hslopes@cpgei.cefetpr.br, [3]alex@ppgia.pucpr.br

Abstract: *This work presents a classification algorithm based on genetic algorithms (GAs) that discovers comprehensible IF-THEN rules, in the spirit of data mining. The proposed GA has a flexible chromosome encoding where each chromosome corresponds to a classification rule. Although the number of genes (genotype) is fixed, the number of rule conditions (phenotype) is variable. The GA also has specific mutation operators for this chromosome encoding. The algorithm was evaluated on two public domain, real-world data sets (on the medical domains of dermatology and breast cancer).*

Keywords: *genetic algorithms, data mining, breast cancer, dermatology, knowledge discovery.*

1 Introdução

Este trabalho apresenta um sistema baseado em algoritmos genéticos (AG) para executar a tarefa de classificação. O sistema foi avaliado em dois domínios médicos: diagnóstico de doenças dermatológicas e previsão de reincidência de câncer de mama. O uso de AG em classificação é uma tentativa de explorar eficientemente o espaço de busca associado a esta tarefa, que normalmente é muito grande. O AG proposto foi projetado de acordo com os conceitos de *Data Mining* (DM) e *Knowledge Discovery*, onde o objetivo não é apenas encontrar conhecimento preciso, mas também conhecimento compreensível [6,8]. Consequentemente, os indivíduos do Algoritmo Genético (AG) representam regras de classificação do tipo SE-ENTÃO, de modo semelhante àquelas descobertas por algoritmos de DM baseados no paradigma de indução de regras.

Este artigo é organizado do seguinte modo, a seção 2 revisa brevemente as características de AG e a tarefa de classificação (do ponto de vista de DM). A seção 3 descreve em detalhes o sistema desenvolvido. A seção 4 descreve as bases de dados usadas nos experimentos. A seção 5 discute os resultados obtidos. A seção 6 apresenta alguns trabalhos relacionados e compara suas principais características com a presente abordagem. Finalmente, na seção 7 são apresentadas as conclusões.

2 Uma Visão Geral de Classificação e AG

A tarefa de classificação é uma das mais estudadas em DM. Em essência, o problema se resume na classificação de registros em classes, através da descoberta de algum tipo de relacionamento entre os atributos. Cada registro (daqui em diante chamado de exemplo) consiste de um conjunto de atributos ditos previsores e um atributo denominado meta [8,11]. Um algoritmo de DM é aplicado a um conjunto de exemplos de treinamento com uma classe conhecida, para descobrir regras que detectem algum relacionamento entre os atributos previsores e o atributo meta. Esta relação é então usada para prever a classe dos exemplos cuja classe é desconhecida, ou seja, para prever o valor do atributo meta.

O conhecimento descoberto é normalmente representado sob forma de regras de previsão do tipo SE-ENTÃO, que têm a vantagem de ser uma representação de conhecimento de alto nível, contribuindo para a compreensibilidade do conhecimento descoberto. As regras descobertas podem ser avaliadas de acordo com vários critérios, tais como: grau de confiança desejado na previsão, taxa de acerto de classificação de exemplos de classe desconhecida, compreensibilidade, etc. Neste artigo será enfatizada a compreensibilidade, a qual é um critério crucial no contexto de DM.

Algoritmos genéticos (AG) são métodos de busca que têm sido extensamente usados em aplicações onde o espaço de busca é muito grande. Em essência, um AG é "um algoritmo de busca baseado nos mecanismos da seleção natural e na genética" [9]. AGs são inspirados no princípio de sobrevivência do melhor indivíduo, ou seja, aquele de melhor *fitness*, os quais têm maior probabilidade de sobreviver e produzir descendência para a geração seguinte. No contexto de busca, os indivíduos são soluções candidatas para um problema específico. Consequentemente a reprodução dos melhores indivíduos significa a reprodução das melhores soluções candidatas correntes. Operadores genéticos como seleção, cruzamento e mutação são os responsáveis pela geração da descendência dos melhores indivíduos. Uma das vantagens de AG em relação aos métodos de busca tradicionais é que o AG

executa uma busca global, usando uma população de indivíduos ao invés de executar uma busca local do tipo *hill-climbing*. Métodos de busca global são preferidos por serem menos sujeitos a máximos locais quando comparados a métodos de busca locais.

Deve ser observado que o paradigma mais usado para a descoberta de conhecimento em DM ainda é a indução de regras e que a maioria dos algoritmos neste paradigma executa algum tipo de busca local.

3 O Algoritmo Genético

O AG proposto neste trabalho foi desenvolvido baseando-se no sistema GALOPPS versão 3.2 [10]. Esta é uma ferramenta de domínio público que incorpora várias características proposta por pesquisadores de AG e que possui grande portabilidade. Nas próximas subseções serão detalhadas as características principais do AG.

3.1 Codificação dos Indivíduos

Um cromossomo é dividido em n genes, onde cada gene corresponde a uma condição envolvendo um atributo, e n é o número de atributos previsores dos dados analisados. Os genes são posicionais, ou seja, o primeiro gene representa o primeiro atributo, o segundo gene representa o segundo atributo, e assim por diante. Cada *i-ésimo* gene, $i=1...n$, é subdividido em três campos: PESO (P_i), OPERADOR (O_i) e VALOR (V_i), como mostrado na figura 1. Cada gene corresponde a uma condição da parte SE de uma regra, e o cromossomo inteiro corresponde a parte SE completa da regra. A parte ENTÃO não precisa ser codificada no cromossomo como será explicado adiante. Portanto, daqui em diante a parte SE da regra codificada será referenciada apenas como regra. Cada indivíduo é representado por apenas um cromossomo. Assim, neste trabalho será utilizado o termo indivíduo como sinônimo de cromossomo.

Gene $_1$			…	…	Gene$_n$		
P_1	O_1	V_1	…	…	P_n	O_n	V_n

Figura 1. Representação de um cromossomo (indivíduo)

O campo PESO (P_i) é uma variável real que pode ter valores entre [0..1]. Esta variável indica se o atributo correspondente está ou não presente na regra. Mais precisamente, quando (P_i) é menor que o valor definido pelo usuário (chamado LIMITE) a *i-ésima* condição da regra é removida. Então, definindo-se um maior valor para LIMITE, aumenta-se a probabilidade da condição correspondente estar presente na regra. No experimento foi usado LIMITE=0,3. As operações de mutação do campo PESO são as responsáveis pela remoção ou inserção do atributo correspondente na regra.

O campo OPERADOR (O_i) é uma variável que indica o operador relacional empregado na *i-ésima* condição. Se o atributo A_i é categórico (nominal), este pode conter os operadores "=" ou "≠", se for contínuo, pode conter os operadores "≥" ou "<".

O campo VALOR (V_i) contém um dos valores permitidos no domínio do atributo A_i. O campo V_i é codificado em um *string* binário, onde o número de *bits* usados para codificar V_i é proporcional ao número de valores possíveis para o atributo A_i. A codificação adotada é bastante flexível com respeito ao comprimento das regras, pois um AG "tradicional" é muito limitado neste aspecto permitindo apenas a manipulação de regras de tamanho fixo. Nesta proposta, embora cada cromossomo tenha um comprimento fixo, os genes são "interpretados" de tal modo (dependendo do valor do campo PESO) que a expressão fenotípica do indivíduo (regra) possa ter um comprimento variável.

3.2 Operadores genéticos

Para os experimentos foram utilizados os operadores genéticos convencionais de seleção (torneio estocástico de tamanho 3) e cruzamento (de dois pontos com probabilidade de 100%). Foi usada uma estratégia de reprodução elitista onde o melhor indivíduo de cada geração é passado inalterado à geração seguinte.

Foram desenvolvidos três operadores de mutação adequados para a representação genética proposta, são eles, mutação de peso, mutação de operador relacional e mutação de valor. Cada um destes operadores age no campo correspondente da representação genética adotada (seção 3.1). A taxa de mutação empregada foi de 30% para cada uma.

A mutação de peso visa modificar o campo PESO de cada condição de regra. Este operador gera um número real pequeno que é então somado ou diminuído do valor corrente no campo PESO, sendo respeitados os limites inferior e superior da variável. Consequentemente, condições de uma regra podem ser removidas ou inseridas, de acordo com o valor de PESO que pode ser maior ou menor que o valor da variável LIMITE, definida pelo usuário.

A mutação de operador relacional modifica o operador relacional que está sendo usado em uma condição da regra, substituindo-o por outro aleatoriamente gerado entre os operadores relacionais válidos (dependendo do atributo, se é categórico ou contínuo).

A mutação de valor visa substituir o conteúdo do campo VALOR por outro aleatoriamente gerado. Há dois casos possíveis. Primeiro, se o atributo é categórico, esta mutação substitui o conteúdo atual do campo por outro que pertença ao domínio do atributo. Segundo, se o atributo é contínuo, gera-se um número pequeno que é somado ou subtraído no conteúdo atual

do campo. Isto é feito de tal modo que não sejam excedidos os limites inferior e superior do domínio do atributo.

3.3 Função de *Fitness*

A função de *fitness* avalia a qualidade de cada regra (indivíduo). Este trabalho usa a função de *fitness* proposta por [14]. Antes de definir a função, deve-se recordar alguns conceitos de avaliação de regras de classificação. Quando é usada uma regra para classificar um paciente, dependendo da classe prevista pela regra e da classe verdadeira do exemplo, podem ser observados quatro tipos diferentes de resultados para a previsão, como segue:

- verdadeiro positivo (*vp*) - a regra prediz que o paciente tem uma determinada doença e o paciente realmente tem a doença;
- falso positivo (*fp*) - a regra prediz que o paciente tem uma determinada doença mas o paciente não tem a doença;
- verdadeiro negativo (*vn*) - a regra prediz que o paciente não tem uma determinada doença, e realmente o paciente não tem;
- falso negativo (*fn*) - a regra prediz que o paciente não tem uma determinada doença mas o paciente tem.

A função de *fitness* combina dois indicadores comuns em domínios médicos, a Sensibilidade (*S*) e a Especificidade (*E*), definidos nas equações 1 e 2.

$$S = vp / (vp + fn) \quad (1)$$

$$E = vn / (vn + fp) \quad (2)$$

Finalmente, a função de *fitness* usada no sistema proposto é definida como o produto destes dois indicadores. Por isso, o objetivo do sistema é maximizar ao mesmo tempo *S* e *E*, e o produto mostrado na equação 3 fornece um bom gradiente para a função.

$$fitness = S * E \quad (3)$$

Cada execução do AG resolve um problema de classificação de duas classes onde a meta é prever se o paciente tem ou não uma determinada doença. Então, o AG é executado pelo menos uma vez para cada classe (valor do atributo de meta). Por exemplo, supondo que o domínio da aplicação tenha seis classes, é preciso executar o AG seis vezes. Na primeira execução o AG busca regras que prevêem a classe 1; na segunda execução busca regras para a classe 2, e assim por diante. Quando o AG está buscando regras que prevêem uma determinada classe, todas as outras classes são fundidas em uma única grande classe que pode ser entendida conceitualmente como significando que o paciente não tem a doença prevista pela regra. Assim, podem ser aplicadas as equações 1 e 2 a problemas com qualquer número de classes.

Esta característica do AG também explica a razão de não ser necessário codificar a classe prevista por uma regra na representação do cromossomo, como mencionado na seção 3.1. Em uma determinada execução do AG todos os indivíduos estão procurando regras que predizem a mesma classe.

4 Bases de Dados Usadas nos Experimentos

Foram feitos experimentos com duas bases de dados de domínio público, nos domínios médicos de dermatologia e câncer de mama.

Estes domínios foram obtidos do UCI (*University of California at Irvine*) - *Machine Learning Repository* [16]. Estes domínios são extensivamente usados em tarefas de classificação que utilizam paradigmas diferentes, ver [3] e [5]. As principais características de cada um destes domínios são tratadas adiante.

4.1 Base de Dados de Dermatologia

O diagnóstico diferencial da doença *erythemato-squamous* é um problema importante em dermatologia. Há seis diagnósticos diferentes (seis classes), e todos eles possuem algumas características clínicas de *erythema* e *scaling*, com poucas diferenças. As seis classes são: *psoriasis*, *seboreic dermatitis*, *lichen planus*, *pityriasis rosea*, *chronic dermatitis* e *pityriasis rubra pilaris*. Algumas características são mais incidentes em certas doenças, mas elas também podem aparecer em algumas fases de desenvolvimento das outras doenças o que pode tornar mais difícil o diagnóstico. Outras características das doenças são discutidas em [5].

Esta base de dados contém 366 registros, cada um com 34 atributos. Todos os atributos têm seus valores mapeados em uma escala de [0..3], onde 0 indica a ausência da característica correspondente e 3 indica uma grande incidência da característica. Duas exceções são o atributo *age* que tem seus valores expressos em anos, e o atributo *family history* que assume o valor 1 quando a doença foi observada na família e 0 caso contrário.

4.2 Base de Dados de Câncer de Mama

O câncer de mama é reincidente em até 30% das pacientes que sofrem uma cirurgia de câncer de mama.

Esta base de dados contém 286 registros, cada um com 9 atributos, e a meta é determinar os pacientes que terão reincidência do câncer de mama. Portanto, há

apenas duas classes, isto é *no-recurrence -events* e *recurrence-events*. Todos os atributos são categóricos. Informações mais detalhadas sobre esta base de dados pode ser encontrada em [3].

5 Resultados Computacionais

Cada execução do AG usou uma população de 50 indivíduos e 50 gerações. O conjunto de parâmetros usados pelos operadores genéticos (cruzamento e mutação, etc.) foram apresentados na seção 3.

Cada base de dados foi dividida aleatoriamente em duas partes, com 2/3 dos registros usados para treinamento e 1/3 dos registros usados para teste da qualidade das regras descobertas. Como sugerido na literatura, esta divisão foi feita de tal modo que a proporção de exemplos que pertencem a cada classe (a freqüência relativa da classe) fosse mantida nos conjuntos de treinamento e de teste. Alguns registros de cada base de dados possuem atributos ausentes e a solução adotada foi remover os poucos registros que continham algum atributo ausente.

5.1 Resultados para a Base de Dados de Dermatologia

A tabela 1 apresenta as seis regras finais descobertas pelo AG - uma regra para cada classe. O AG foi executado três vezes para cada classe variando-se a semente aleatória de geração da população inicial. A melhor regra das três execuções, de acordo com seu valor de fitness medido no conjunto de treinamento, foi selecionada como a regra previsora daquela classe (esta é a regra apresentada na tabela 1). Uma vez que as seis regras foram selecionadas, elas foram avaliadas usando os dados de teste. Note-se que o conjunto de regras usado para classificação também inclui uma regra "padrão", ou seja, uma regra sem nenhuma condição que é automaticamente aplicada quando nenhuma outra regra tem suas condições satisfeitas pelo exemplo que está sendo classificado. Esta regra simplesmente prevê classe 1 que é a "classe da maioria", isto é, a classe mais freqüente nos dados de treinamento.

Para cada regra da tabela 1 são mostrados dois valores que são o *fitness* da regra computada pela equação 3 nos dados de treinamento e nos dados de teste respectivamente. Pode-se ver que todas as regras descobertas nos dados de treinamento generalizam bem as amostras nos dados de teste. Na maioria dos casos o *fitness* nos dados de teste é quase igual ao *fitness* dos dados de treinamento. As únicas exceções são as regra para as classes C3 e C4, onde o *fitness* encontrado nos dados de treinamento não é mantido nos dados de teste.

Os valores de *fitness* informados na tabela 1 são úteis para avaliar o desempenho de cada regra individualmente. Porém, também é importante avaliar o desempenho do conjunto de regras como um todo. Como é usual na literatura, esta avaliação foi feita através da taxa de acerto nos dados de teste, ou seja, a razão entre o número de exemplos corretamente classificados e o número total de exemplos nos dados de teste. A taxa de acerto do conjunto de regras foi de 95% (dos 119 exemplos, 113 foram classificados corretamente).

Tabela 1. Regras descobertas para a Base de Dados de Dermatologia

C	Regra	*Fitness*
1	**SE** (*clubbing of the rete ridges*)≥1 **E** (*perifollicular parakeratosis*)=0	0,973-0,973
2	**SE** (*koebner phenomenon*)=0 **E** (*vacuolisation and damage of basal layer*)<1 **E** (*spongiosis*)≥2	0,855-0,855
3	**SE** (*band-like infiltrate*)≥2	1,000-0,979
4	**SE** (*knee and elbow involvement*)=0 **E** (*family history*)=0 **E** (*acanthosis*)<3 **E** (*focal hypergranulosis*)<2 **E** (*spongiosis*)≥1 **E** (*inflammatory monoluclear infiltrate*)≥1	0,860-0,783
5	**SE** (*melanin incontinence*)<1 **E** (*fibrosis of the papillary dermis*)≠0 **E** (*munro microabcess*)=0	1,000-1,000
6	**SE** (*follicular papules*)≥1 **E** (*perifollicular parakeratosis*)≥1	1,000-1,000

Também foi observado que em 5 das 6 classes, o conjunto de regras descobertas era similar nas três execuções para cada classe. Considerando que diferentes sementes foram utilizadas para a geração da população inicial, pode-se concluir que o algoritmo convergiu e identificou regras confiáveis para cada classe a partir dos atributos fornecidos. Entretanto, essa conclusão precisa ser confirmada através de novos experimentos.

5.2 Resultados para a Base de Dados de Câncer de Mama

A tabela 2 mostra as duas regras finais descobertas pelo AG - uma regra para cada classe. Os procedimentos adotados foram os mesmos descritos na seção 5.1, executando o AG três vezes para cada classe.

Para cada regra da tabela 2 são mostrados dois valores que são o *fitness* da regra computada pela equação 3 nos dados de treinamento e nos dados de teste respectivamente. Desta vez, entretanto, as regras descobertas a partir dos dados de treinamento não generalizaram muito bem os exemplos dos dados de teste. Isto parece ser devido ao fato de que este é um

problema de classificação consideravelmente mais difícil em comparação com a base de dados de dermatologia, e um outro agravante é o ruído presente nesta base de dados.

De qualquer maneira, é interessante avaliar o desempenho do conjunto de regras descobertas como um todo, medindo a taxa de acerto, como feito na seção anterior. A taxa de acerto do conjunto de regras descobertas foi 67% (dos 92 exemplos, foram classificados corretamente 62).

Tabela 2. Regras descobertas para a Base de Dados de Câncer de Mama

C	Regra	*Fitness*
1	**SE** (*inv-nodes*)<(9-11) **E** (*deg-malig*)<2 **E** (*irradiat*)=(no)	0,564-0,365
2	**SE** (*menopause*)≥(ge40) **E** (*tumor-size*)≠(45-49) **E** (*deg-malig*)≥2	0,497-0,393

De maneira análoga aos resultados de seção 5.1, para cada classe foram descobertas regras compostas pelas mesmas condições nas três execuções do AG, mostrando que o AG converge para a mesma melhor regra apesar da variação da semente inicial.

6 Trabalhos Relacionados

Vários AGs têm sido projetados e propostos na literatura para descobrir algum tipo de regra de classificação compreensível. Alguns destes trabalhos são brevemente descritos abaixo.

GIL [12] usa vários operadores de generalização/especialização propostos por [15] para estender os operadores genéticos de AG convencionais, criando um AG de conhecimento-intensivo para a tarefa de classificação. GIL segue a abordagem de Pittsburgh para aprendizado de regras, onde cada indivíduo da população corresponde a um conjunto de regras. GABIL [4] também segue a abordagem de Pittsburgh e também sugere alguns operadores genéticos específicos para a tarefa. Ainda outro projeto que segue a abordagem de Pittsburgh é HDPDCS [18]. Este sistema foi projetado especificamente para classificação e extração de características de bases de dados de alta dimensionalidade.

Uma diferença principal entre o sistema proposto e os projetos citados é que no nosso sistema não foi seguida a abordagem de Pittsburgh. Neste sistema, um indivíduo corresponde a uma única regra, o que tende a ser computacionalmente mais eficiente. Esta abordagem é normalmente conhecida como abordagem de Michigan.

O sistema REGAL aprende regras de lógica de primeira ordem (*first-order-logic*) - FOL [17]. A descoberta de regras FOL é uma característica interessante compartilhada por poucos AG, mas no caso de REGAL, ela requer que o usuário forneça um tipo de modelo da fórmula lógica a ser aprendida. Esta exigência reduz a autonomia do sistema, o que pode ser considerada uma desvantagem no contexto de DM. Alguns AGs paralelos são descendentes do REGAL, por exemplo o sistema G-NET [1].

SIAO1 [2] também aprende descrições de classe FOL, mas não precisa de um modelo fornecido pelo usuário. Ao invés disso, ele adapta a técnica de generalizar um exemplo como semente para aprender regras de classificação FOL [15]. Do ponto de vista de DM, parece que a desvantagem principal de SIAO1 é que este tipo de técnica de generalização é computacionalmente cara, o que torna o SIAO1 difícil de aplicar a bancos de dados grandes.

Ambos REGAL e SIAO1 são muito diferentes do sistema proposto e um sistema um pouco mais semelhante é o EDRL [13]. Este sistema também requer uma execução do AG para descobrir uma regra para cada classe. Também, neste sistema cada cromossomo é representado como uma conjunção de condições de uma única regra, similarmente ao sistema proposto. Porém, este sistema não pode tratar com atributos contínuos, isto é, assume que os atributos contínuos foram previamente discretizados como um passo de pré-processamento. O sistema proposto não tem esta limitação, isto é, trabalha com atributos categóricos e contínuos.

Outro AG para descobrir regras que segue a abordagem de associar um cromossomo com uma conjunção de condições de uma única regra é descrito em [7]. Porém, este sistema não é direcionado à tarefa de classificação. Ele é direcionado à tarefa de modelagem de dependências (ou indução de regras generalizadas). Esta tarefa pode ser considerada como uma generalização da anterior, pois regras diferentes podem prever atributos meta diferentes (classes), ao invés de um único atributo meta.

7 Conclusões e Trabalhos Futuros

Os resultados preliminares apresentados neste artigo são promissores, e permitem concluir que a codificação do cromossomo e a representação do conjunto de regras associada é uma boa alternativa para extrair regras compreensíveis no contexto de DM.

As taxas de precisão alcançadas pelo AG proposto na base de dados de dermatologia (95%) e na base de dados de câncer de mama (67%) são semelhantes às encontradas por outros pesquisadores. Porém, o AG parece ser particularmente efetivo para encontrar um conjunto conciso de regras compreensíveis, considerando que ele descobre uma única regra para cada classe. Outros algoritmos de DM

descobrem freqüentemente várias regras para uma única classe, o que pode tornar difícil para o usuário entender as diversas regras descobertas.

Um trabalho futuro deveria consistir de mais experiências com outras bases de dados, bem como, outras experiências para aperfeiçoar os vários parâmetros do algoritmo, como taxas de mutação e o limiar da variável LIMITE usada pelo campo PESO. Os resultados alcançados não consideraram avaliações mais precisas sobre os valores mais adequados para os parâmetros adotados

Agradecimentos

O domínio de câncer de mama foi obtido do Centro Médico Universitário, Instituto de Oncologia, Ljubljana, Iugoslávia. Agradecimentos para M. Zwitter e M. Soklic por disponibilizarem os dados.

Agradecimentos a H. A. Guvenir da Universidade de Bilkent e N. Ilter da Universidade de Gazi, Sukara, Turquia, por disponibilizar o domínio de dermatologia.

Referências Bibliográficas

[1] Anglano, C.; Giordana, A.; Lo Bello, G.; Saitta, L. A network genetic algorithm for concept learning. *Proc. 7th Int. Conf. Genetic Algorithms*, p. 434-441, Morgan Kaufmann, 1997.

[2] Augier, S.; Venturini, G.; Kodratoff, Y. Learning first order logic rules with a genetic algorithm. *Proc. 1st Int. Conf. Knowledge Discovery & Data Mining*, p. 21-26. Montreal, Canada, 1995.

[3] Clark, P., Niblett, T. Induction in Noisy Domains. In: *Progress in Machine Learning* (from the Proceedings of the 2nd European Working Session on Learning), p. 11-30, Sigma Press, 1987.

[4] DeJong, K.A.; Spears, W.M.; Gordon, D.F. Using genetic algorithms for concept learning. *Machine Learning*, v. 13, p. 161-188, 1993.

[5] Demiroz, G.; Govenir, H.A.; Ilter N. Learning differential diagnosis of erythemato-squamous diseases using voting feature. *Artificial Intelligence in Medicine*, v. 13, p. 147-165, 1998.

[6] Fayyad, V.M.; Piatetsky-Shapiro, C.; Smyth, P. From data mining to knowledge discovery: na overview. Fayyad, V.M. et al. (Eds.) *Advances in knowledge Discovery and Data Mining*, p. 1-34, AAAI/MIT Press, 1996.

[7] Freitas, A.A. A genetic algorithm for generalized rule induction. R. Roy et al. (Eds.) *Advances in Soft Computing – Engineering Design and Manufacturing*, p. 340-353. Springer-Verlag, 1999.

[8] Freitas, A.A.; Lavington, S.H. *Mining Very Large Databases with Parallel Processing*, London: Kluwer Academic Publishers, 1998.

[9] Goldberg, D.E. *Genetic Algorithms in Search, Optimization and Machine Learning*. Reading: Addison-Wesley, 1989.

[10] Goodman, E.D. *An Introduction to GALLOPS - The Genetic Algorithms Optimized for Portability and Parallelism System*. East Lansing, USA: Genetic Algorithms Research and Applications Group (GARAGe), Department of Computer Science, Michigan State University, 1996.

[11] Hand, D. *Construction and Assessment if Classification Rules*. Chichester: John Wiley & Sons, 1997.

[12] Janikow, C.Z. A knowledge-intensive genetic algorithm for supervised learning. *Machine Learning*, v. 13, p. 189-228, 1993.

[13] Kwedlo, W.; Kretowski, M. Discovering of decision rules from databases: an evolutionary approach. *Principles of Data Mining and Knowledge Discovery, Proc. 2nd European Symp*. Lecture Notes in Artificial Intelligence 1510, p. 370-378.

[14] Lopes, H.S.; Coutinho, M.S.; Lima, W.C. An evolutionary approach to simulate cognitive feedback learning in medical domain. In: Sanchez, E., Shibata, T., Zadeh, L.A. (Eds) *Genetic Algorithms and Fuzzy Logic Systems*. Singapore: World Scientific, p.193-207, 1997.

[15] Michalski, R.W. A theory and methodology of inductive learning. *Artificial Intelligence*, v. 20, p. 111-161, 1983.

[16] Murphy, P. M.; Aha, D. W. UCI Repository of Machine Learning databases. [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, Departament of Information and Computer Science, 1994.

[17] Neri, F.; Giordana, A. A parallel genetic algorithm for concept learning. *Proc. 6th Int. Conf. Genetic Algorithms*, p. 436-443, 1995.

[18] Pei, M.; Goodman, E.D.; Punch III, W.F. Pattern discovery from data using genetic algorithms. *Proc. 1st Pacific-Asia Conf. Knowledge Discovery and Data Mining*, 1997.