

3

BAS Genético: Uma abordagem evolutiva

O algoritmo BAS possui uma prova de corretude para a utilização de uma distribuição inicial arbitrária para os exemplos de treinamento. No entanto, encontrar a melhor distribuição inicial dado um conjunto de exemplos de treinamento continua sendo um problema em aberto.

Neste capítulo, apresentamos uma heurística para o problema de determinação da melhor distribuição inicial para o BAS. Tal heurística evolucionária baseia-se em uma estratégia amplamente utilizada, os Algoritmos Genéticos.

3.1

Algoritmos Genéticos

Algoritmos Genéticos (AG) são uma família de modelos computacionais inspirados nos processos da Evolução e Seleção Natural (Hol75). Os AG modelam a solução de um problema em uma estrutura chamada **cromossomo** que representa as possíveis soluções, chamadas de **indivíduos**. Uma série de operadores genéticos é aplicada em tais cromossomos com o objetivo de otimizar a solução.

Dois componentes têm uma função importante no método: a codificação do problema e a função de aptidão.

A *codificação do problema* é o mapeamento realizado entre os cromossomos e os indivíduos. Normalmente, os indivíduos são mapeados de duas formas possíveis, a saber:

1. Cadeia de 1s e 0s, indicando a presença ou não de uma característica;
2. Cadeia de valores inteiros, indicando possíveis valores para cada característica do indivíduo.

Exemplos dos dois mapeamentos mais comuns são mostrados na Figura 3.1.

A *função de aptidão* visa a quantificação da otimalidade de um cromossomo de forma a comparar o seu desempenho como solução do problema com o de outros indivíduos. Cromossomos ótimos, ou pelo menos “mais” ótimos,

Características	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
Mapeamento 1	1	1	1	0	0	1	1	0	1
Mapeamento 2	5	3	3	6	1	0	2	5	10

Figura 3.1: Exemplos de cromossomos.

são os escolhidos para serem aplicados pelos diversos operadores genéticos, gerando uma nova população que, possivelmente, será ainda melhor.

O algoritmo genético inicia com uma população aleatória de indivíduos. A população corrente é transformada pelos operadores genéticos em uma nova geração de indivíduos. O objetivo principal da nova geração é o de manter os *melhores* indivíduos, segundo as suas aptidões particulares, aumentando a aptidão global da população, até que um critério de parada seja atingido. Tal critério pode variar entre limiares de aptidão, número de gerações ou falta de melhora.

Existem dois tipos de operadores genéticos básicos: seleção e recombinação. Operadores de seleção utilizam-se da função de aptidão para decidir quais indivíduos tem o maior potencial. Tais indivíduos devem persistir na população seguinte e também devem ser utilizados pelo outro tipo de operador.

Os operadores de recombinação são utilizados para criar novos indivíduos utilizando um ou mais indivíduos cujo potencial de aptidão seja alto. Seu objetivo é diversificar o processo de busca. Os operadores desse tipo mais famosos são o cruzamento e a mutação.

O operador de cruzamento utiliza duas ou mais frações de indivíduos com alto potencial de aptidão com o objetivo de criar um novo indivíduo que é adicionado à próxima geração da população.

O operador de mutação, por outro lado, utiliza um indivíduo com alto potencial de aptidão e realiza uma pequena alteração em um de seus componentes. O novo indivíduo também é acrescentado à próxima geração da população.

Um possível pseudo-código de alto nível para o algoritmo genético é dado no Algoritmo 3.1.

3.2 Modelagem Genética Utilizada

Nesta seção, mostramos a codificação genética utilizada de forma a obter uma boa distribuição inicial para os exemplos de treinamento do algoritmo BAS. Primeiramente, o conjunto de dados de treinamento original é dividido em dois conjuntos: treinamento genético e validação.

Algoritmo 3.1 Algoritmo Genético.

-
- 1: Escolha uma população aleatória de indivíduos
 - 2: Avalie a aptidão dos indivíduos
 - 3: **repita**
 - 4: Selecione os *melhores* indivíduos a serem utilizados pelos operadores genéticos
 - 5: Gere novos indivíduos utilizando cruzamento e mutação
 - 6: Avalie a aptidão dos novos indivíduos
 - 7: Troque os *piores* indivíduos da população pelos *melhores* novos indivíduos
 - 8: **até** algum critério de parada
-

3.2.1**Codificação Genética para o BAS**

Na codificação genética desenvolvida para o BAS, o cromossomo, cujo tamanho é igual ao tamanho do conjunto de treinamento genético, é composto pelos possíveis valores de pesos para os exemplos. Os valores de entrada necessários para tal codificação são apenas os limites de valores possíveis para os pesos, {pesoMínimo, pesoMáximo}. Um exemplo de tal codificação é dado na Figura 3.2, mostrando um indivíduo cujo cromossomo tem tamanho 7 e pesos com valores inteiros entre 1 e 5.

	p_1	p_2	p_3	p_4	p_5	p_6	p_7
Cr	3	4	2	1	3	5	2

Figura 3.2: Exemplo da codificação genética para o BAS.

3.2.2**Função de aptidão**

Utilizando o conjunto de treinamento genético, um classificador BAS é treinado e associado a cada indivíduo da população. Uma medida de desempenho, como por exemplo a acurácia, do classificador BAS gerado no conjunto de validação é utilizada como valor de aptidão para cada indivíduo para todas as gerações.

3.2.3**Operador de cruzamento**

O operador de cruzamento gera um novo cromossomo, quebrando dois cromossomos em um ponto aleatório e os combinando. A Figura 3.3 mostra um exemplo de operador de cruzamento para a abordagem BAS Genética proposta.

	p_1	p_2	p_3	p_4	p_5	p_6	p_7
Cr_1	3	4	2	1	3	5	2
Cr_2	2	3	5	2	1	4	3
$Cr_1 \otimes Cr_2$	3	4	2	2	1	4	3

Figura 3.3: Exemplo do operador de cruzamento.

3.2.4 Operador de Mutação

O operador de mutação gera um novo cromossomo alterando o valor de um dos pesos dos exemplos para um outro valor de peso possível. A Figura 3.4 mostra um exemplo de operador de mutação para a abordagem BAS Genética proposta.

	p_1	p_2	p_3	p_4	p_5	p_6	p_7
Cr_1	3	4	2	1	3	5	2
$\odot Cr_1$	3	4	2	5	3	5	2

Figura 3.4: Exemplo do operador de mutação.

3.2.5 Escolha da melhor distribuição

O algoritmo mantém uma cópia do melhor indivíduo de todas as gerações do algoritmo genético, baseado em sua aptidão. Como a distribuição de pesos do melhor indivíduo somente cobre os exemplos utilizados no treinamento genético, o conjunto de validação não pode ser utilizado no treinamento do classificador final e é removido do conjunto de dados originais. Com essa distribuição final, um classificador BAS é treinado utilizando tal conjunto de dados modificado.