

PROGRAMAÇÃO GENÉTICA COM INICIALIZAÇÃO BASEADA EM FLORESTA RANDÔMICA EM ESTUDOS DE ASSOCIAÇÃO DO GENOMA COMPLETO

IGOR MAGALHÃES RIBEIRO¹; CARLOS CRISTIANO HASENCLEVER BORGES¹; WAGNER ANTONIO ARBEX²; BRUNO ZONOVELLI DA SILVA¹
1 – UNIVERSIDADE FEDERAL DE JUIZ DE FORA; 2 – EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA;
wagner.arbex@embrapa.br

Resumo – Interações epistáticas entre (SNPs) são responsáveis por doenças complexas como alguns tipos de câncer. Estudos de associação do genoma completo objetivam encontrar tais interações no genótipo de uma população para inferir os fatores de risco ou a predisposição do indivíduo a determinado fenótipo. Algumas doenças como câncer cervical, leucemia e diabetes do tipo 2 apresentam baixa herdabilidade, o que torna mais difícil a descoberta destas associações. Este artigo apresenta um algoritmo de programação genética com inicialização baseada em floresta randômica para identificar interações (SNP-SNP) quando há baixa herdabilidade. O objetivo é selecionar os SNPs mais importantes e introduzi-los na população inicial da PG. Bases de dados do tipo caso-controle foram simuladas com o software GAMETES variando a herdabilidade em 0,4, 0,3, 0,2 e 0,1. Os experimentos comparam o algoritmo com e sem o método de inicialização e os resultados mostram que a utilização da floresta randômica aumenta consideravelmente as chances de descoberta das interações entre os SNPs causais.

Palavras-chave: Inteligência Computacional. Programação Genética. Floresta Randômica. (GWAS). Bioinformática.

I. INTRODUÇÃO

Os avanços da tecnologia de genotipagem de DNA em massa permitem que um grande número de variações de sequências sejam medidas ao longo do genoma. Diversas plataformas comerciais como Illumina (BENNET, 2004), AppliedBiosystems (JANITZ, 2008) e Affymetrix têm a capacidade de analisar mais de um milhão de SNPs sobre um chip.

Um polimorfismo de base única (do inglês, *single nucleotide polymorphism* - SNP) é uma variação em uma única posição na sequência de DNA entre indivíduos de uma mesma espécie. SNPs ocorrem aproximadamente, a cada 600 bases e aparecem por todo o código genético. Essas variações correspondem a alternância (substituição, ausência ou duplicação) dos nucleotídeos A, T, C e G em uma frequência alélica mínima de 1% em uma determinada população (BROOKES, 1999) (GUIMARÃES; COSTA, 2002).

SNPs são utilizados como marcadores genéticos e para fins computacionais, estas plataformas de genotipagem geram dados na forma de letras (ex: aa, Aa, AA) ou números (ex: 0, 1, 2) que definem os alelos encontrados em cada indivíduo (OLAZAR, 2013).

A partir dos marcadores SNPs, obtém-se a representação do genótipo de cada indivíduo. O fenótipo pode ser explicado como sendo qualquer característica

mensurável, podendo o mesmo ser discreto ou contínuo e é resultado da interação entre o genótipo de o ambiente (ex: cor dos olhos, cabelo ou alguma doença específica).

Em estudos de associação do genoma completo (do inglês, genome wide association studies - GWAS) utiliza-se os dados do genótipo para tentar inferir o fenótipo. GWAS utiliza todo o genoma para procurar associações entre marcadores com determinadas doenças. O objetivo principal deste tipo de estudo é encontrar a partir destas associações, novos candidatos para as variantes dos genes causadores da doença (fenótipo). Em outras palavras, a ideia é encontrar os mecanismos genéticos de doenças e os seus fatores de risco.

Este trabalho utiliza-se de estudos de caso-controle. Queremos determinar quais dos diversos SNPs presentes na amostra são responsáveis pela manifestação ou não da característica de interesse. Para tal, o fenótipo é dado na forma de 1 (caso) ou 0 (controle).

Casos e controles devem ser de grupos étnicos similares para evitar diferenças genéticas sutis que enviesariam os resultados do estudo. Neste tipo de estudo, o DNA de ambos os grupos de pessoas é utilizado. A Figura 1 representa a esquematização de um estudo de caso-controle em GWAS.

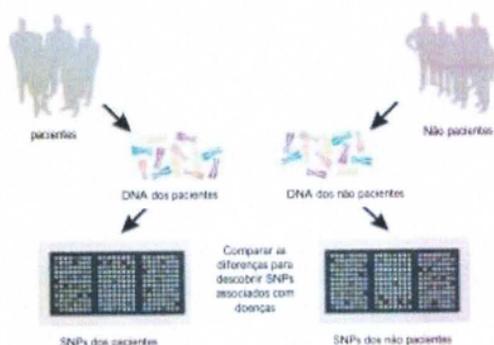
Doenças complexas podem ser causadas por interações gênicas epistáticas. A ação de epistasia é causada por interações entre os genes. A epistasia é inferida quando um alelo de um gene altera a expressão padrão de outros genes. O gene que exerce a ação inibitória é denominado de epistático e o que têm seu efeito mascarado é chamado de hipostático (AJF MILLER JH, 2000).

Como exemplo de epistasia, suponha que existam dois genes A e B. Sejam os fenótipos $Bb_{aa} = 8$ e $bb_{aa} = 6$. Supondo que o a seguinte alteração genotípica de troca do alelo 'a' pelo alelo 'A' e produza as seguintes alterações no fenótipo anterior: $Bb_{Aa} = 12$ e $bb_{Aa} = 4$. Desta forma, no primeiro fenótipo pode-se observar que a substituição resultou em uma variação positiva igual a 4. No segundo, a substituição resultou em uma supressão e a variação foi negativa igual a -2. Outro ponto a ser analisado é a herdabilidade.

A herdabilidade pode ser estimada pela razão entre as variâncias do genótipo e fenótipo. Esta razão mede a proporcionalidade de quanto o fator genético influencia sobre o fenótipo (ex: se um fenótipo ou doença apresenta herdabilidade de 0,4, isso indicia que 40% do total da

variação fenotípica da doença é devido a constituição genética).

Figura 1 - A associação entre cada SNPs e o fenótipo é calculada com base na prevalência de cada SNPs nos pacientes (casos) e nos não pacientes (controles).



Fonte: (OLAZAR, 2013).

Diversas condições médicas ou doenças apresentam baixa herdabilidade como por exemplo Asma (0,3), câncer de bexiga (0,07-0,31), câncer cervical (22), leucemia (0,01) (CZNESE;LICHTENSTEIN;HEMMINKI, 2002), câncer de pulmão (0,08), hipertensão (0,3), doença de Parkinson (0,25-0,3) (DO et al., 2011), diabetes do tipo-2 (0,26) (POULSEN;KYVIK;BECK-NIELSEN, 1999). Faz-se então necessário desenvolver algoritmos capazes de identificar fatores de risco em diferentes níveis de herdabilidade.

A herdabilidade presente na amostra interfere diretamente na capacidade de classificação correta entre caso-controle. Quanto menor a herdabilidade, menor é a explicação obtida através do genótipo, e de forma inversa maior a influência dos fatores ambientais.

O objetivo deste trabalho é desenvolver um algoritmo capaz de identificar associações epistáticas entre o genótipo e a fenótipo em diferentes níveis de herdabilidade. Uma técnica promissora é a programação genética

II. PROGRAMAÇÃO GENÉTICA

A programação genética é uma meta-heurística proposta por Koza (KOZA, 1992) pertencente a classe dos algoritmos evolucionários. O seu objetivo é evoluir programas de computador (indivíduos) em uma população por meio de operadores genéticos (reprodução e mutação) que levam a possíveis soluções desejadas a partir de um conjunto de entrada.

A ideia básica de um algoritmo de programação genética pode ser vista na Figura 2. Inicialmente uma população com um determinado número de indivíduos é criada para dar início a um processo de evolução iterativo. Esses indivíduos são avaliados e em cada geração, novos indivíduos mais promissores que os seus progenitores são criados e inseridos na população por meio dos operadores genéticos de cruzamento e mutação. Dado o seguinte fluxograma de um algoritmo de programação genética, os seguintes passos são executados (AUGUSTO, 2009) :

1. Criar população de indivíduos: No início do processo iterativo, a população inicial é criada a partir de indivíduos aleatórios. Antes dessa etapa devem ser definidos a representação dos indivíduos e os conjuntos de funções e terminais que definem a saída esperada da solução.

2. Avaliar os novos indivíduos: A cada geração os indivíduos da população devem ser avaliados para determinar a sua proximidade com a saída desejada.

3. Critério de parada: caso o critério de parada seja satisfeito, o algoritmo deve interromper a sua execução. Normalmente utiliza-se dois tipos de critérios de parada:

(a) quando a solução ótima é encontrada.

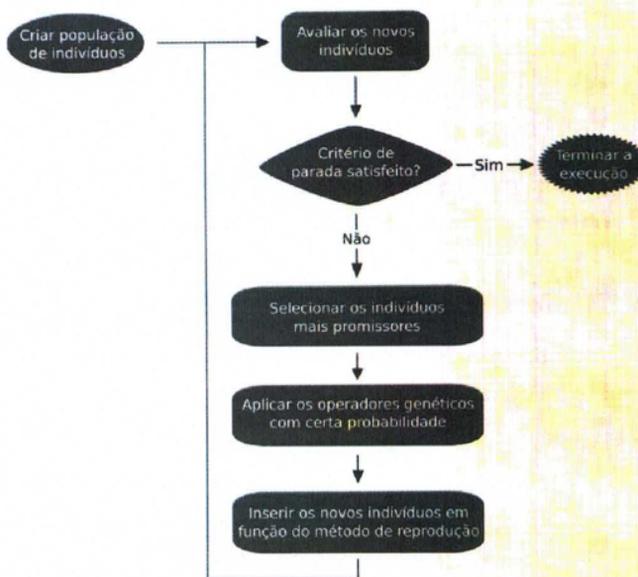
(b) quando o número de gerações limite é atingido. Neste último caso, o melhor indivíduo encontrado ao longo das gerações é o escolhido como a solução do problema.

4. Selecionar os indivíduos mais promissores: os indivíduos que obtêm os melhores resultados de avaliação tem maior probabilidade de gerarem proles mais eficientes para a resolução do problema.

5. Aplicar os operadores genéticos com certa probabilidade: Operados genéticos como reprodução e mutação são aplicados aos indivíduos selecionados para gerar diversidade na população.

6. Os novos indivíduos gerados são adicionados na população através de um dos métodos de reprodução.

Figura 2 - Fluxograma de um algoritmo de programação genética.



Fonte: Extraído de (AUGUSTO, 2009).

III. MOTIVAÇÃO

A programação genética foi utilizada em GWAS nos seguintes trabalhos (NUNKESSER, 2007), (SZE-To *et al.*, 2013) e (ESTRADA-GIL *et al.*, 2007) para a geração de modelos discriminativos.

A vantagem dos modelos discriminativos é que eles possuem a capacidade de aprender modelos de classificação diretamente. Desta forma podem gerar regras para prever se indivíduos podem ser susceptíveis ou não. As regras geradas neste processo podem ser interpretadas e um especialista pode decidir se a saída do algoritmo é condizente com o problema biológico e se elas podem ser utilizadas em aplicações reais.

Estes trabalhos adotam modelos de árvore de decisão para representar as soluções da PG. Entretanto, dos três citados somente o GP-Pi utiliza um mecanismo de

inicialização da população inicial para guiar a evolução das soluções da PG. Moore e White (2007) sugeriram que apesar da programação genética ser melhor do que uma busca randômica para este tipo de problema. Sem uma inicialização da população inicial apropriada o problema se torna mais complexo com o aumento no número de SNPs.

IV. BASE DE DADOS

As bases de dados utilizadas no trabalho foram simuladas com o *software* GAMETES (URBANOWICZ, 2012) que simula epistasia entre SNPs. O GAMETES gera bases de dados de marcadores SNPs bi-alélicos com o objetivo de simular modelos de doenças complexas. O Software permite variar os valores de herdabilidade, menor frequência alélica, tamanho da população e quantidade de marcadores.

Figura 3 - Exemplo de uma base de dados gerada pelo software GAMETES, as n primeiras colunas representam o genótipo dado pelos SNPs e a última coluna representa o fenótipo (0 para controle e 1 para caso).

SNP1	SNP2	SNP3	SNP4	SNP5	SNP6	SNP7	SNP8	SNP9	SNP10	F
0	0	0	0	1	1	1	0	0	0	0
0	0	0	0	1	2	0	0	1	0	1
1	0	0	0	0	0	1	1	1	1	0
0	0	2	0	0	1	0	0	0	0	0
1	0	1	0	0	1	0	0	0	0	1
0	0	0	0	0	0	1	0	1	2	1
1	1	0	0	1	2	0	0	0	2	0
0	0	1	0	0	1	0	0	1	1	1
2	0	0	0	2	0	0	0	0	0	0
1	1	0	0	0	0	0	0	1	2	1
2	0	0	0	1	0	0	0	0	0	0
0	0	0	0	2	1	0	0	0	1	1

Fonte: Autores, 2016.

Neste trabalho foram simulados quatro conjuntos de dados. Como objetivo é desenvolver um algoritmo de PG capaz de identificar corretamente os SNPs causais em diferentes níveis de herdabilidade, somente este parâmetro foi alterado. Com exceção da herdabilidade, os demais parâmetros foram fixados de acordo com a Tabela 1, os valores definidos foram obtidos do trabalho de Sze-To *et al.* (2013).

Tabela 1 – Parâmetros utilizados na simulação das bases de dados com o *software* GAMETES.

Parâmetros	Valor
Menor frequência alélica	0,2
População (P)	2000
Número de SNPs	100
Herdabilidade	0,1; 0,2; 0,3; 0,4

A saída do software GAMETES consiste em uma matriz $P \times SNPs + 1$, onde a última coluna na matriz representa a classe de cada indivíduo (fenótipo). O genótipo é dado pelas colunas de SNPs e são representados por: 0, 1 ou 2 onde 0 é identificado como homocigoto dominante (AA), 1 é o heterocigoto (Aa) e 2 é o homocigoto recessivo (aa). A Figura 3 mostra um exemplo da matriz resultante da simulação.

V. PROCEDIMENTOS

Neste trabalho, foi desenvolvido um algoritmo de programação genética e um método eficiente para

inicialização da população inicial. O método é capaz de encontrar interações SNP-SNP mesmo em casos onde a herdabilidade é baixa. A junção desses métodos possibilita que os SNPs causais tenham uma maior probabilidade de participarem do processo evolutivo.

O método de inicialização, basicamente, consiste em utilizar o *rank* da floresta randômica. Para a construção do algoritmo de programação genética foi utilizada uma biblioteca desenvolvida em Luke (2007). A probabilidade de crossover adotada foi de 0,9, a probabilidade de mutação foi de 0,05 e a máxima profundidade das árvores que representam os indivíduos foi de 6. Para o conjunto de terminais foram utilizados dois elementos (=, !=) e para o conjunto de funções (E, OU). Todos os parâmetros podem ser encontrados na Tabela 2. Os valores dos parâmetros foram baseados em Sze-To *et al.* (2013).

Tabela 2 – Parâmetros do algoritmo de PG.

Parâmetros	Valor
Gerações	50
Tamanho População	4096
Tipo de cruzamento	Um ponto
Taxa de cruzamento	0,9
Taxa de mutação	0,05
Método de seleção	Sete torneios
Conjunto de funções	E, Ou
Conjunto de terminais	!=, =
Profundidade máxima das árvores	6

A representação dos indivíduos é baseada em (NUNKESSER, 2007) e é utilizada para expressar as interações entre SNPs na forma normal disjuntiva (do inglês, *disjunctive normal form* - DNF). Uma expressão lógica em DNF é uma disjunção de um ou mais monômios, onde um monômio é constituído de um único literal ou um conjunto dos mesmos. A Figura 4 demonstra uma árvore genérica com expressões lógicas em DNF representando um indivíduo da PG. A gramática utilizada é simples e o conjunto de funções é dado pelas expressões "E" e "Ou" enquanto que o conjunto de terminais é definido pelos SNPs e sua respectivas saídas, por exemplo "SNP 1 = 0".

4.1 – Inicialização da população inicial do algoritmo de programação genética

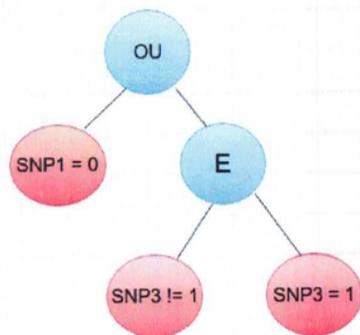
O mecanismo de inicialização da população inicial foi baseado em floresta randômica (BREIMAN, 2001). O objetivo é garantir que os possíveis SNPs causais tenham mais chance de aparecer na população inicial.

A floresta randômica é um modelo de classificação que combina o resultado de diversos classificadores, sendo que cada classificador corresponde a uma árvore de decisão. A classificação final corresponde ao voto da maioria das árvores. Neste trabalho foi utilizado o pacote *randomForest* escrito em R (todos os parâmetros foram *default*, exceto o número de árvores de decisão utilizadas pelo modelo que foi igual a 1500), o pacote está disponível em <<https://cran.r-project.org/web/packages/randomForest/>>.

O método foi utilizado para ranquear os SNPs. Antes do início do processo da PG, os SNPs são avaliados pela floresta randômica. Logo, quando um indivíduo é criado na população inicial, é realizada uma seleção por 7 torneios para decidir o SNP que irá popular cada nodo da árvore que representa o indivíduo. Dos 7, o que for melhor ranqueado pela floresta randômica é escolhido. O processo é realizado

até que os nodos terminais do indivíduo sejam completamente preenchidos. Cada SNP pode aparecer somente uma vez por árvore.

Figura 4 - Representação de um indivíduo da PG na forma DNF. Os nodos do conjunto de funções estão representados em azul e os nodos do conjunto de terminais estão em vermelho e representam os SNPs e sua variação alélica.



Fonte: Adaptada de (NUNKESSER, 2007).

Trabalhos como GP-Pi utilizam o algoritmo de seleção de atributos ReliefF para iterativamente estimar os pesos dos SNPs baseados em padrões de indivíduos vizinhos. Uma variação do algoritmo também foi utilizado como inicializador em (MOORE; WHITE, 2007) e representa uma melhora do modelo inicial, onde utiliza os k vizinhos mais próximos para estimação do peso.

4.2 – Função de Avaliação

A função de avaliação do algoritmo de programação genética é dada pela Equação (1)

$$f_i = (P / (VP + VN)) + N_i / \alpha \quad (1)$$

e representa uma variação da função de avaliação desenvolvida em (NUNKESSER, 2007). P representa o número total de indivíduos casos-controle. VP é o número de verdadeiros positivos e VN os verdadeiros negativos classificados corretamente. N_i representa o número de nodos do indivíduo da programação genética e α é uma constante de parcimônia, para os experimentos foi utilizado o valor de 6. A expressão privilegia os resultados onde as soluções são compostas por árvores menores.

V. RESULTADOS

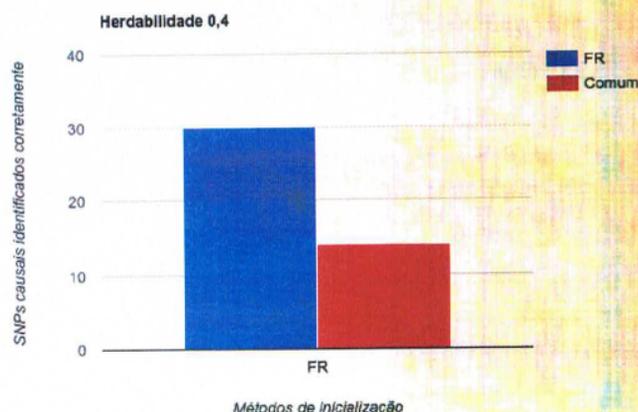
Os experimentos apresentados foram realizados sobre bases de dados simulados conforme os parâmetros apresentados na Tabela 1.

Para cada método, a herdabilidade variou entre 0,4 - 0,1 sendo executadas 30 rodadas para cada variação do algoritmo. Todas as simulações foram feitas com 100 SNPs, onde dois são considerados funcionais (SNP99 e SNP100). Os resultados obtidos, com e sem a utilização do mecanismo de inicialização, são apresentados a seguir.

A herdabilidade define, de forma genérica, a influencia genética sobre o fenótipo. O valor de 0,4 é um valor considerado razoável para esse parâmetro, de forma que a identificação dos SNPs causais pode ser facilitada. A classificação correta dos casos-controles é diretamente impactada pela herdabilidade. A inicialização da população inicial usando floresta randômica visa minimizar esse problema.

A Figura 5 apresenta os resultados da simulação com herdabilidade igual a 0,4. Pode-se observar que neste caso, o método sem a inicialização proposta consegue encontrar os SNPs funcionais em quase metade das execuções do algoritmo de PG. Entretanto, com a utilização do mecanismo proposto, em todas as 30 execuções os SNPs funcionais foram encontrados.

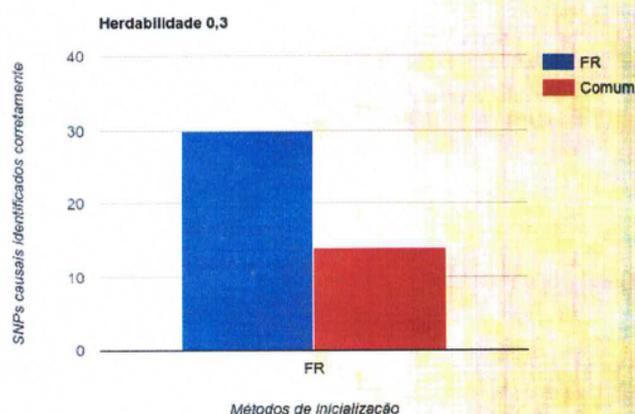
Figura 5 – Base de dados simulada com herdabilidade = 0,4



Fonte: Autores, 2016.

Os resultados dos experimentos obtidos com a base de dados com herdabilidade = 0,3 se assemelham com os anteriores (Figura 6). Pode-se observar que sem o método de inicialização, em quase metade das execuções os SNPs causais foram encontrados.

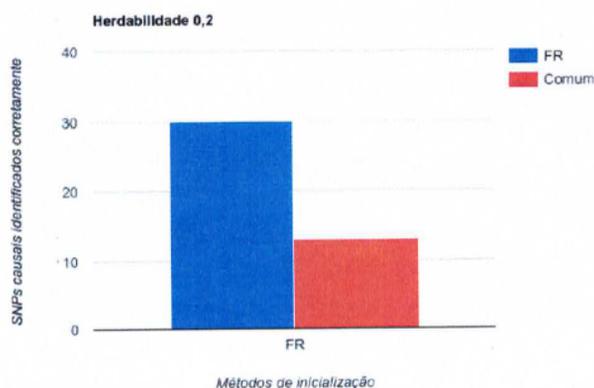
Figura 6 – Base de dados simulada com herdabilidade = 0,3



Fonte: Autores, 2016.

O mesmo ocorreu quando a herdabilidade = 0,2, (13 execuções), como pode ser visto na Figura 7. O que indica que até esse nível de herdabilidade, em pouco mais do que um terço dos casos, o algoritmo sem o mecanismo de inicialização proposto ainda conseguiu identificar os SNPs causais.

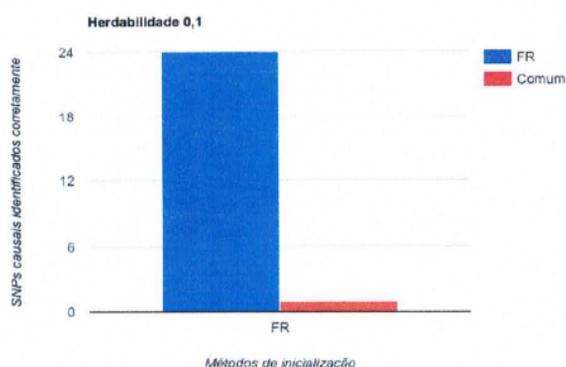
Figura 7 – Base de dados simulada com herdabilidade = 0,2.



Fonte: Autores, 2016.

Com o nível de herdabilidade em 0,1 (Figura 8), pode-se notar a necessidade de um mecanismo de inicialização da população inicial já que sem o seu uso, nas 30 execuções, os SNPs funcionais foram detectados apenas em uma das execuções. Para o caso mais complexo, com herdabilidade 0,1, pode-se notar que em algumas execuções os SNPs causais não foram encontrados mesmo com a utilização do mecanismo de inicialização.

Figura 8 - Base de dados simulada com herdabilidade = 0,1



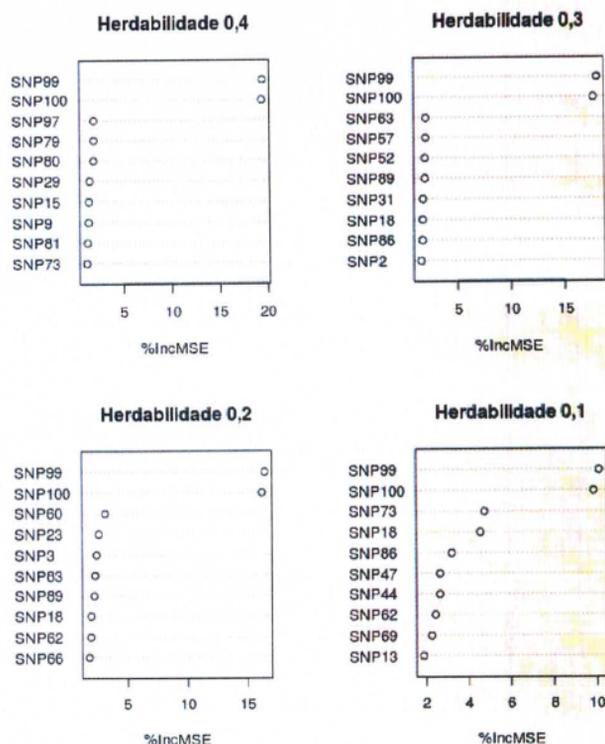
Fonte: Autores, 2016.

A Figura 9 mostra o aumento percentual do erro quadrático médio dos dez melhores SNPs resultantes do ranqueamento da floresta randômica nas bases de dados simuladas. O resultado do aumento percentual do erro quadrático médio por ser robusto e informativo, permite uma análise direta dos resultados. A floresta randômica também oferece como saída o índice de Gini que em problemas de classificação, como os desse trabalho, pode vir a ser uma medida mais interessante, sendo inclusive tema de trabalhos futuros. Quando a herdabilidade é 0,3 ou 0,4 o método consegue facilmente selecionar os SNPs causais.

A Figura 9 mostra o *rank* da RF com herdabilidade de 0,2 e 0,1. O *rank*, com herdabilidade de 0,2, mostra um gráfico entre 0 e 15 com os SNPs causais próximos de 15 e os não causais abaixo de 5. Entretanto, com herdabilidade de 0,1 o gráfico vai até somente 10 e os não causais estão acima de 6. Logo, pode-se observar que a medida que o

nível de herdabilidade diminui, ambos os SNPs causais e não causais tendem a se aproximar. Entretanto, a floresta randômica ainda consegue selecionar os SNPs mais importantes mesmo em casos onde o problema se torna mais complexo, fazendo com que esses marcadores tenham mais chances de aparecer na população inicial e segregarem durante o processo evolutivo da PG.

Figura 9 – Rank da RF.



Fonte: Autores, 2016.

VI. CONCLUSÃO

Neste artigo, apresentou-se um algoritmo de programação genética capaz de descobrir interações SNP-SNP com baixa herdabilidade através de um inicializador da população inicial. O inicializador proposto explora os SNPs obtidos através do rank da floresta randômica (FR). Os experimentos foram executados em bases de dados simuladas com o software GAMETES e os resultados dos experimentos que utiliza e que não utiliza o inicializador foram comparados.

Pode-se observar que os resultados em dados simulados mostram que o modelo proposto pode ser uma alternativa aos modelos de inicialização existentes na literatura e que o mesmo pode ajudar a alcançar as regras de interação mesmo quando a herdabilidade é baixa, demonstrando que o nosso método pode melhorar o processo de busca pelos SNPs causais.

Como trabalhos futuros, outras medidas de ordenação da floresta randômica devem ser analisadas. Também devem ser realizados experimentos com bases de dados reais e com números de SNPs maiores utilizando o modelo proposto.

VII. REFERÊNCIAS BIBLIOGRÁFICAS

AJF MILLER JH, S. D. e. a. G. An Introduction to Genetic Analysis. 7nd. ed. New York:W. H. Freeman, 2000.

Disponível em:

<<http://www.ncbi.nlm.nih.gov/books/NBK21766/>>.

AUGUSTO, D. Programação Genética Multi-populacional e Co-evolucionária para classificação de Dados. Tese (Doutorado) — COPPE/UFRJ, 8 2009.

BENNETT, S. Solexa Ltd. **Pharmacogenomics**, v.5, n.4, p.433–438, Junho 2004. ISSN 1462-2416

BREIMAN, L. Random forests. **Machine Learning**, v.45,p.5–32, Outubro 2001.

BROOKES, A. J. The essence of snps. **Gene**, v. 234, n. 2, p. 177 – 186, 1999. ISSN 0378-1119. Disponível em:<<http://www.sciencedirect.com/science/article/pii/S037811199900219X>>

CZENE, K., LICHTENSTEIN, P., HEMMINKI K. Environmental and heritable causes of cancer among 9.6 million individuals in the Swedish Family-Cancer Database. **Int J Cancer**. Maio ,2002.

DO, C. B., et al. Web-Based Genome-Wide Association Study Identifies Two Novel Loci and a Substantial Genetic Component for Parkinson's Disease.**PLoS Genetics**, 2011.

ESTRADA-GIL, J. et al. Gpdti: A genetic programming decision tree induction method to find epistatic effects in common complex diseases. **Bioinformatics**, v.23, n.13,p. i167–i174, 2007.

GUIMARÃES, P.; COSTA, M. Snps: sutis diferenças de um código. **Biotechnol. Cienc. Desenvolv**, v. 26, p. 24–27, 2002.

JANITZ, M. Next-Generation Genome Sequencing: Towards Personalized Medicine . **Wiley**, 2008. ISBN 9783527320905.

KOZA, J. Genetic Programming: On the Programming of Computers by Means of Natural Selection. **Bradford**, 1992. (A Bradford book). ISBN 9780262111706.

LUKE, S., et al. Ecj: A java-based evolutionary computation research system, 2007.

MOORE, J., WHITE, B. Genome-wide genetic analysis using genetic programming: The critical need for expert knowledge. **Genetic Programming Theory and Practice IV**, 2007.

NUNKESSER, R. et al. Detecting high-order interactions of single nucleotide polymorphisms using genetic programming. **Bioinformatics**, v.23,n.24,p.3280–3288, 2007. Disponível em: <http://bioinformatics.oxfordjournals.org/content/23/24/328>

OLAZAR, M. Uma Metodologia para a Descoberta de Marcadores Genéticos em Estudos de Associação. Tese (Doutorado) – COPPE/UFRJ, 5 2013.

POULSEN, P. et al. Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. **Diabetologia**. v.42,n.2,p.139–145. Fevereiro 1999.

SZE-TO, H.-Y., et al. {GP-Pi}: Using Genetic Programming with Penalization and Initialization on Genome-Wide Association Study. **Artificial Intelligence and Soft Computing - 12th International Conference, {ICAISC} 2013**, Proceedings, Part {II}, v.7895,p.330–34,1 Zakopane, Poland, Junho 2013.

URBANOWICZ, R. J. et al. METHODOLOGY GAMETES : a fast , direct algorithm for generating pure , strict , epistatic models with random architectures. **BioData Mining 2012**, 5:16, v.5,n.16,p.1–14. 2012.

VIII. COPYRIGHT

Direitos autorais: O(s) autor(es) é(são) o(s) único(s) responsável(is) pelo material incluído no artigo.

REVISTA



SOLUÇÕES PARA O DESENVOLVIMENTO DO PAÍS

Atendimento:
sodebras@sodebras.com.br
Acesso:
<http://www.sodebras.com.br>

REFLEXÕES SOBRE ARTE E DESIGN NA SOCIEDADE DE CONSUMO – Marco Antonio Rossi; Elaine Patricia Grandini Serrano	79
ANALYSIS OF THE VARIABILITY IN THE POSITION OF MEASUREMENT IN OFFENSIVE GOALBALL – Altemir Trapp; Alessandro Tosim; Maria Lucia Miyake Okumura; Osiris Canciglieri Junior; Marcelo Rudek	84
SISTEMAS DA INFORMAÇÃO NO ESPORTE: IMPLICAÇÕES PARA O GOALBALL – Altemir Trapp; Alessandro Tosim; Maria Lucia Miyake Okumura; Osiris Canciglieri Junior; Marcelo Rudek	88
ADUBAÇÃO NITROGENADA DA CULTURA DO TRIGO COM BASE NA CLOROFILOMETRIA VIA REMOTELY-PILOTED AIRCRAFT – Marcos Antonio Moretto; Cristiano Reschke Lajús; Gean Lopes Da Luz; Fernando Chiesa; Neomar Sandrin	92
PROSPECÇÃO DE PATENTES RELACIONADAS AO USO DE AERONAVE REMOTAMENTE PILOTADA COMO INSERÇÃO TECNOLÓGICA APLICADA EM AGRICULTURA DE PRECISÃO – Marcos Antonio Moretto; Giovani Echer; Cristiano Reschke Lajús; Gean Lopes Da Luz; Dhouglas Ricardo Pedruzzi	97
PROGRAMAÇÃO GENÉTICA COM INICIALIZAÇÃO BASEADA EM FLORESTA RANDÔMICA EM ESTUDOS DE ASSOCIAÇÃO DO GENOMA COMPLETO – Igor Magalhães Ribeiro; Carlos Cristiano Hasenclever Borges; Wagner Antonio Arbex; Bruno Zonovelli Da Silva	104
REGRESSÃO COM MÁQUINAS DE VETORES SUPORTE E SELEÇÃO DE ATRIBUTOS VIA ALGORITMO GENÉTICO APLICADA EM SELEÇÃO GENÔMICA – Bruno Zonovelli; Carlos Cristiano Hasenclever Borges; Wagner Antonio Arbex; Fabrizzio Condé De Oliveira; Igor Magalhães Ribeiro	110
SEPARAÇÃO DAS REGIÕES DE CÉU E TERRA EM IMAGENS DIGITAIS – Arlete Teresinha Beuren; Jacques Facon	116
A DINÂMICA DOS PROCESSOS RELACIONADOS ÀS INUNDAÇÕES NO MUNICÍPIO DE ITAPERUNA-RJ – Juliana Ribeiro Costa; Antonio Ferreira Da Hora	122
CARACTERIZAÇÃO DOS INDICADORES AMBIENTAIS EM UM EMPREENDIMENTO DA CONSTRUÇÃO CIVIL – ESTUDO DE CASO – Anai De Lima Nogueira; Ronaldo Pimentel Mannarino; Joecila Santos Da Silva	126
INFLUENCIA DA ADIÇÃO DE NIÓBIO EM AÇOS UTILIZADOS NA FABRICAÇÃO DE MOLAS AUTOMOTIVAS – Heitor Barbosa Soldatti; Valdir Alves Guimarães; Daniela Helena Pelegrine	132
A UTILIZAÇÃO DE ANGULARJS E SIGNALR EM SISTEMA SUPERVISÓRIO – Wilian Douglas Dos Santos Penaforte; Henrique Glicério Da Conceição Gomes; Fábio De Paula Carvalho; Demétrio Renó Magalhães; Mateus Sales André Cruz; Silvano Fonseca Paganoto; Célia De Jesus Vidal	138
CONCEITUAÇÃO DO PROTOCOLO AUTONAV – Demétrio Renó Magalhães; Silvano Fonseca Paganoto; Henrique G. Da Conceição Gomes; Wilian Douglas Dos Santos Penaforte; Celia De Jesus Vidal; Fábio De Paula Carvalho	141
GANHO DE PRODUTIVIDADE EM DOCUMENTAÇÃO DE SOFTWARE COM O ENTERPRISE ARCHITECT – Fábio De Paula Carvalho; Demétrio Renó Magalhães; Silvano Fonseca Paganoto; Wilian Douglas Dos Santos Penaforte; Henrique G. Da Conceição Gomes; Célia De Jesus Vidal	146
INTERFACE HUMANO-MÁQUINA PARA WORLD WIDE WEB – Henrique Glicério Da Conceição Gomes; Wilian Douglas Dos Santos Penaforte; Fábio De Paula Carvalho; Demétrio Renó Magalhães; Silvano Fonseca Paganoto; Célia De Jesus Vidal	151
MODELO DE PREVISÃO DO CONSUMO DE ELETRICIDADE EM UM EDIFÍCIO EDUCACIONAL – Abreu, Jacksiel, J. E; Cavalcante, C. A. M. T.	154
USO DE AQUECEDORES SOLARES DE GARRAFA PET PARA PISCINAS – Evaldo Chagas Gouvêa, Teófilo Miguel De Souza	159
ANÁLISE DO CUSTO BENEFÍCIO DAS FERRAMENTAS DE CRIMPAGEM DE DIFERENTES FORNECEDORES – Camila De Jesus Rodrigues; Juliana Anhaia De Oliveira; Luis Carlos Machado	165