

UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA
DISCIPLINA DE COMPUTAÇÃO EVOLUTIVA
RELATÓRIO DE PESQUISA – ALGORITMO EVOLUCIONÁRIO

Título do Projeto: Notebooks Avell

Pesquisador: Leonardo de Lima Souza

Orientador: Danilo Araújo

1. Introdução

O presente relatório detalha o desenvolvimento, análise e aplicação de um Algoritmo Evolucionário (AE) na tentativa de minimizar a dificuldade para escolher e comprar um notebook da marca Avell no site desta empresa, levando em consideração que o site além de oferecer modelos pré configurados também disponibiliza uma ferramenta de personalização dos mesmos.

Embora a personalização dos equipamentos simbolize uma vantagem para o cliente, que poderá escolher os principais componentes, também torna-se um desafio, pois é possível montar mais de 130.000 notebooks diferentes a partir das opções de personalização disponíveis, sendo estas: processador, tipo e tamanho de armazenamento, memória RAM, placa vídeo, sistema operacional entre outras.

A Avell é uma empresa brasileira fundada no ano 2000 e tem atuação principal no Brasil e Estados Unidos. De acordo com o site da empresa, ela foca-se somente na fabricação e comercialização de notebooks, sendo estes diferenciados por serem equipamentos de alto desempenho e muito personalizáveis.

De forma geral notebooks de alto desempenho, que podem ser utilizados em diversos contextos como jogos ou uso profissional em manipulação de mídias de alta resolução ou tratamento de grandes massas de dados, têm consequentemente um valor de compra/venda também alto, o que torna o nicho de mercado para esses equipamentos um pouco restrito.

A proposta deste trabalho é utilizar um AE para encontrar quais são as melhores opções de notebooks da marca Avell levando em consideração seu

desempenho e preço, ou seja, um AE multiobjetivo para encontrar notebooks com o melhor desempenho e menor preço possível.

2. Formalização do problema

O objetivo geral do AE desenvolvido é fornecer soluções de notebooks que dispõem de alto desempenho, porém com o menor preço de compra possível, caracterizando assim um problema multiobjetivo.

Acessando o site da empresa (avell.com.br) nota-se disponibilidade de aproximadamente 30 modelos de notebook para venda, todavia esse número pode variar dependendo da situação do mercado, quantidade de vendas e lançamentos de novos modelos.

Os modelos estão dispostos em duas categorias: para jogos e para uso profissional. Em ambas é possível adquirir dispositivos pré configurados bem como o usuário pode montar a especificação de desejada de acordo com os componentes disponíveis para um determinado modelo, isto é, nem todos os componentes estão disponíveis para um certo modelo.

Entre os componentes que podem ser personalizados destacam-se: o processador, a placa de vídeo, a memória RAM, o tipo, tamanho e quantidades de unidades de armazenamento, o sistema operacional e a placa de rede wireless. Uma análise prévia considerando a quantidade de modelos e as várias formas de personalização percebeu-se a possibilidade de montar mais de 130.000 dispositivos diferentes, tornando o processo de escolha de um notebook uma tarefa complexa e dispendiosa para um usuário.

A **Figura 1** exibe uma parte da funcionalidade de personalização onde pode-se escolher por meio de cliques.

Figura 1 - Sistema de personalização

PLACA DE VÍDEO

- ☒ 2x SLI NVIDIA GeForce GTX 1080 (8GB GDDR5X)
- ☐ 1x NVIDIA GeForce GTX 1080 (8GB GDDR5X) (-R\$6.600,00)

PROCESSADOR

- ☒ Intel® Core™ i7-7700 Kaby Lake 3.6 GHz, 8MB Cache (4.2 GHz com Max Turbo)
- ☐ Intel® Core™ i7-7700K Kaby Lake 4.2 GHz, 8MB Cache (4.5 GHz com Max Turbo) (-R\$205,00)

MEMÓRIA

- ☒ 16 GB Memória DDR4 (2400 MHZ)
- ☐ 24 GB Memória DDR4 (2400 MHZ) (-R\$248,00)
- ☐ 32 GB Memória DDR4 (2400 MHZ) (-R\$466,00)
- ☐ 64 GB Memória DDR4 (2400 MHZ) (-R\$1.400,00)
- ☐ 8 GB Memória DDR4 (2133 MHZ) (-R\$248,00)

HARD DISK (HDD), SSD OU SSHD

- ☒ SSHD 1 TB com 8GB SSD
- ☐ SSD 240 GB SATA III
- ☐ HD 2TB - 5400 RPM (-R\$110,00)
- ☐ SSD 480 GB SATA III (-R\$286,00)
- ☐ SSD 960 GB SATA III (-R\$943,00)
- ☐ SATA M.2 - 240GB SSD - [500 Mb/s]
- ☐ SATA M.2 - 480GB SSD - [500 Mb/s] (-R\$286,00)
- ☐ SATA M.2 - 960GB SSD - [500 Mb/s] (-R\$943,00)

Fonte: avell.com.br

Para o desenvolvimento do AE foi necessário montar uma base de dados com as informações dos notebooks disponíveis para venda no site da Avell, um script de coleta foi implementado e todas as informações necessárias foram obtidas em forma de strings e armazenadas em formato Json em um banco de dados Mongo.

Fazendo uma análise prévia dos dados notou-se que estes eram muito despadronizados, ocorrendo inconsistências, como por exemplo, um mesmo componente ser representado por duas strings diferentes em modelos distintos, o que se fez necessário uma normalização e posteriormente migração dos dados resultantes para uma nova base.

O banco padronizado foi dividido em várias coleções, uma delas armazena todos modelos com suas respectivas informações sobre os componentes, as outras coleções são dos componentes armazenados sem registros duplicados. Essas coleções são acessadas no início da execução do AE para definir de forma dinâmica os valores que cada gene pode assumir e variar no cromossomo.

O **Quadro 1** mostra a representação do cromossomo.

Quadro 1

Representação do Cromossomo				
Gene	Componente	Total por componente atual	Limite inferior	Limite superior
0	Nome e modelo	30	0	29
1	Cor	9	0	8
2	Sistema operacional	3	0	2
3	Office	3	0	2
4	Processador	6	0	5
5	Memória RAM	8	0	7
6	Tela	10	0	9
7	1ª unidade de armazenamento	8	0	7
8	2ª unidade de armazenamento	6	0	5
9	1ª unidade SATA eM2	4	0	3
10	2ª unidade SATA eM2	4	0	3
11	Placa de vídeo	10	0	9
12	Placa wireless	7	0	6

Fonte: autor

Na montagem do cromossomo o valor do limite inferior que cada gene pode assumir é sempre 0 (zero), pois este representa o valor de índice de um *array* que é

preenchido a partir dos dados presentes em uma coleção de um componente no banco e de forma análoga o valor limite superior é igual ao tamanho do *array* menos um.

Note-se, por exemplo, no **Quadro 1** que o gente 4 (processador) pode assumir 6 valores diferentes, representando 6 processadores distintos e que o valor do seu limite superior é 5.

2.1. Forma de avaliar o fitness

Para avaliar o fitness de cada solução é considerado o seu preço em reais e uma nota de desempenho variando de 1 a 10. O cálculo dessa nota é realizado por meio das informações presentes no site userbenchmark.com, este dispõe de uma vasta base de dados sobre benchmarks realizados em computadores por usuários localizados globalmente em vários países.

No site encontram-se todos os componentes de hardware (CPU, RAM, HD, SSD, ...) fabricados no últimos anos, avaliados e catalogados de diversas maneiras, como velocidade, preço de compra, avaliação do usuário e etc.. Para o projeto foi considerada a classificação de componentes feita a partir da velocidade.

O **Quadro 2** mostra um exemplo de como ficou a distribuição de pesos para os processadores, que é um dos componentes que influenciam diretamente no desempenho.

Quadro 2

Processadores		
Modelo	Posição userbenchmark	Nota/Peso atribuído
7700K	4	10
7700	21	7.5
7820HK	45	5
7700HQ	90	2.5
7300HQ	109	1

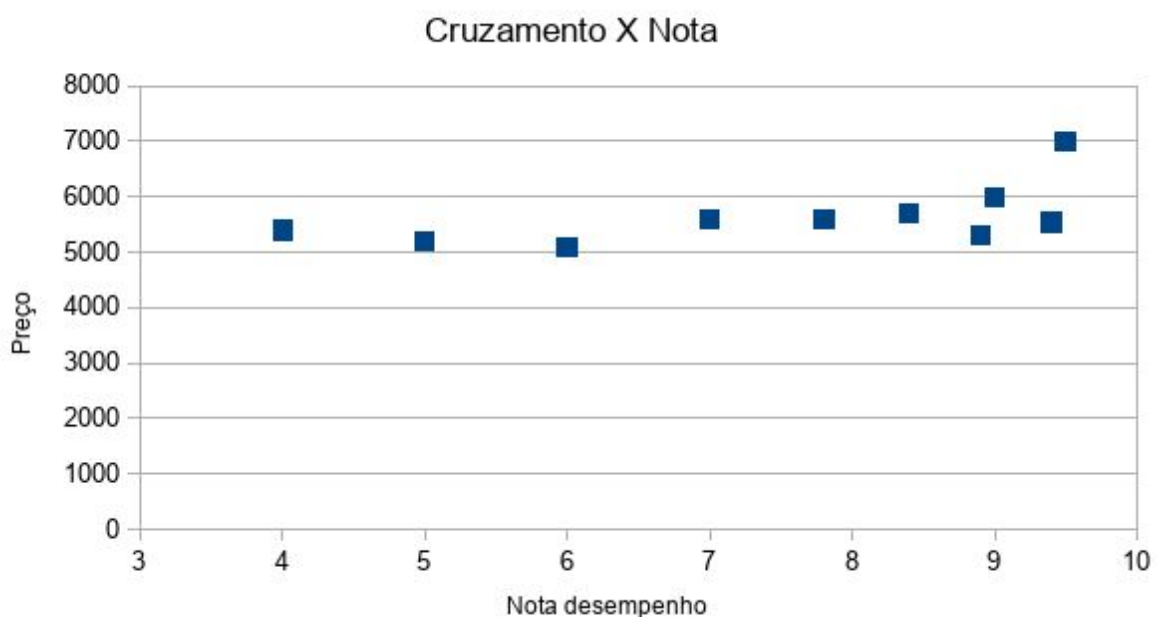
Fonte: autor

A nota final de desempenho é calculada a partir de uma média aritmética simples entre todos os componentes que influenciam na performance do notebook: $(\text{notaProcessador} + \text{notaMemoriaRAM} + \text{notaPrimeiraUnidadeArmazenamento} + \text{notaSegundaUnidadeArmazenamento} + \text{notaPrimeiraSATAeM2} + \text{notaSegundaSATAeM2} + \text{notaPlacaDeVideo}) / 7$. A nota portanto varia de 1 a 10.

No entanto o preço final do equipamento, obtido a partir do somatório dos preços dos seus componentes, também é considerado no cálculo do fitness. Existem duas formas de pagamento com valores diferentes: pagamento no boleto com 10% de desconto ou parcelado em 12 vezes. Neste projeto adota-se o valor do pagamento no boleto.

O fitness é calculado com o cruzamento dos dois valores (nota e preço) em um plano cartesiano, sendo cada notebook representado por um ponto (x = nota, y = preço). Com isso é possível estabelecer o fitness obtendo a Distância Euclidiana entre o ponto que representa o notebook e o ponto (10, 0), onde 10 é o maior valor de nota de desempenho, veja na **Figura 2** o gráfico que ilustra o contexto.

Figura 2



Fonte: autor

O gráfico da **Figura 2** exemplifica a situação onde diversos notebooks estão dispostos no plano e aquele com a menor distância do ponto (10, 0) tem o melhor fitness.

2.2. Penalização

Como os modelos de notebooks não podem ser personalizado livremente com qualquer componente disponível, então ao finalizar o cálculo do fitness é realizada uma verificação para validar se os componentes estão de acordo com o modelo. Caso não estejam, o fitness é penalizado somando-se 0.08 para cada componente inválido referente ao modelo.

3. Metodologia

A variante de AE utilizado no desenvolvimento e execução foi Algoritmo Genético (AG) de forma adaptada para para o contexto do problema que utiliza valores inteiros para representar cada solução. O fluxo de execução de uma AG básico é apresentado a seguir:

1. Inicializa uma população com um fixo de indivíduos e avalia cada um deles definindo o fitness.
2. Escolhe aleatoriamente dois pais para realizar um cruzamento.
3. Executa o cruzamento entre os pais resultando dois filhos, que são salvos em um array intermediário (temporário).

O passo 3 é repetido até que a quantidade de filhos seja igual ao tamanho da população atual.

4. Em cada filho é executada uma mutação com probabilidade de mutação baixa para cada gente da solução.
5. Os filhos são avaliados e o array de pais é substituído pelo array de filhos.

Uns dos desafios em computação evolutiva é a definição de qual é a melhor configuração paramétrica para solucionar o problema em estudo. Isso envolve a escolha de operadores de mutação e recombinação bem como as probabilidades de cada um deles serem executados, tamanho da população e quantidade de gerações por execução.

Para este problema foi escolhido os seguintes parâmetros:

- Cruzamento - 2: crossover de um ponto e crossover de n pontos.
- Mutação - 2: incrementos lentos e inicialização Aleatória.
- Seleção de pais - 2: proporcional ao fitness e seleção uniforme.
- Gerações - 2: 1000 e 5000.
- Tamanho de população - 3: 50; 100; 200.
- Probabilidade de Cruzamento (PC) - 3: 0.75; 0.85; 0.95.
- Probabilidade de Mutação (PM) - 3: 0.03; 0.06; 0.1.

Com isso chega-se a uma quantidade de $2 * 2 * 2 * 2 * 3 * 3 * 3 = 432$ configurações paramétricas diferentes para serem testadas. Para ver uma tabela com essas configurações acesse este link: <https://goo.gl/98adsF>.

Durante os testes cada configuração foi executada 2 vezes, sendo calculados e salvos o menor fitness, maior fitness, menor fitness médio, desvio padrão do fitness médio e o tempo das 2 execuções em segundos. Os resultados das execuções são expostos no tópico 5.

4. Resultados

O quadro 3 exibe as 10 melhores soluções classificadas pelo menor fitness, bem como a configuração paramétrica para alcançar esse resultado e outros dados estatísticos.

Quadro 3

10 melhores execuções											
Gerações	Seleção de pais	Recombinação	Mutação	Tamanho pop	PR	PM	Menor Fitness	Maior Fitness	Menor Fitness Médio	Desvio padrão fitness	Tempo (segundos) 2 execuções
1000	Uniforme	CrossoverNPontos	IncrementosLentos	200	0.75	0.06	4.30	1325.29	21.00	54.38	3.70
5000	FPS	CrossoverNPontos	InicializacaoAleatoria	200	0.95	0.1	4.35	1010.37	5.19	59.48	618.94
5000	Uniforme	CrossoverNPontos	InicializacaoAleatoria	200	0.95	0.06	4.43	993.27	5.34	40.05	17.54
5000	FPS	CrossoverNPontos	IncrementosLentos	100	0.8	0.1	4.69	1499.8	5.65	56.04	152.84

		s			5			9			
5000	Uniforme	CrossoverNPontos	IncrementosLentos	50	0.75	0.1	4.72	2027.44	10.84	54.52	4.76
5000	Uniforme	CrossoverUmPonto	InicializacaoAleatoria	200	0.95	0.06	4.77	2922.70	5.27	80.38	13.20
5000	Uniforme	CrossoverNPontos	InicializacaoAleatoria	100	0.75	0.1	4.80	1365.71	5.61	61.58	7.93
5000	Uniforme	CrossoverNPontos	InicializacaoAleatoria	200	0.75	0.1	4.83	348.93	5.55	26.32	16.01
5000	FPS	CrossoverUmPonto	IncrementosLentos	200	0.85	0.1	4.83	755.44	5.48	42.77	624.20
5000	FPS	CrossoverUmPonto	IncrementosLentos	50	0.75	0.06	4.86	4088.12	7.97	228.24	42.44

Fonte: autor

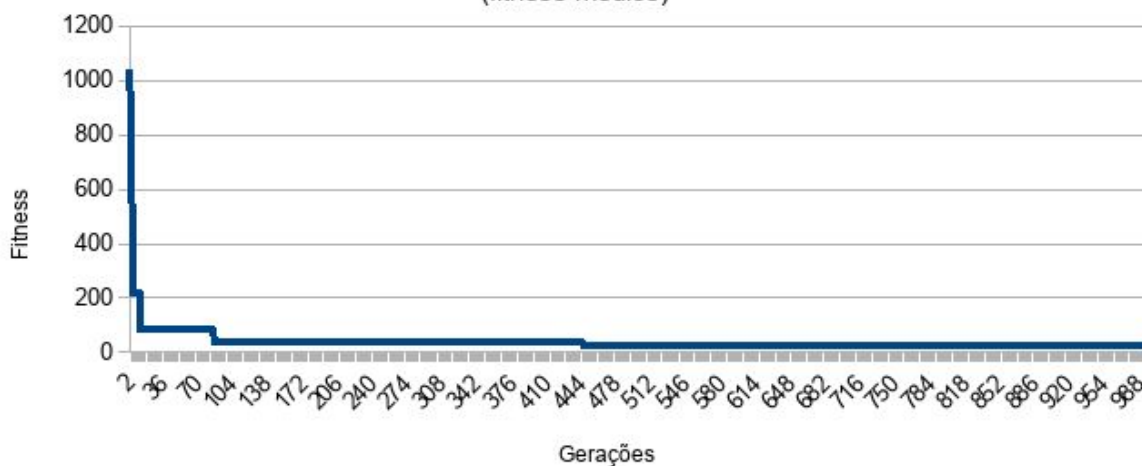
4.1. Resumo da solução número 1

4.1.1. Gráfico de convergência do fitness médio **Figura 3**

Figura 3

Gráfico de convergência Solução 1

(fitness médios)



Fonte: autor

4.1.2. Especificação do notebook da solução 1

Fitness: 4.301769010202859

Especificação do Notebook

Nome do Modeloo = FullRange G1746 IRON V4

Cor

Nome = Azul Fosco

Preço = R\$ 99.0

Sistema Operacional

Descrição = Original Windows 10 Pro 64bits

Preço = R\$ 480.0

Office

Tipo = Microsoft Office 2016 Home & Students (Word, Excel, PowerPoint, OneNote)

Preço = R\$ 399.0

Preço padrão

No boleto = R\$ 9499.5

Cartão de crédito = R\$ 10555.0

Desconto = 10.0%

Parcelas = 12

Preço personalizado

No boleto = R\$ 13285.800000000001

Cartão de crédito = R\$ 14762.0

Desconto = 10.0%

Parcelas = 12

Processador

Fabricante = Intel

Família = i7

Modelo = 7700K

Memória cache = 8

Frequência Máxima = 4.5

Preço = R\$ 205.0

Memória RAM

Tipo = DDR4

Frequência = 2400.0

Tamanho = 32

Preço = R\$ 465.0

Tela

Polegadas = 15.6

Tecnologia = 4K QFHD com Tecnologia G-SYNC (Matte)

Preço = R\$ 590.0

1º Memória de armazenamento

Tipo = SSD SATAe M.2

Unidade de medida = GB

Tamanho = 240

Preço = R\$ 0.0

2º Memória de armazenamento

Tipo = SSD SATA III

Unidade de medida = GB

Tamanho = 240

Preço = R\$ 456.0

1º SATA_eM2 Memória de armazenamento

Tipo = SSD SATAe M.2

Unidade de medida = GB

Tamanho = 480

Preço = R\$ 742.0

2º SATA_eM2 Memória de armazenamento

Tipo = SSD SATAe M.2

Unidade de medida = GB

Tamanho = 480

Preço = R\$ 742.0

Placa de Vídeo

Fabricante = NVIDIA

Modelo = GeForce GTX 1080

Memória dedicada = 8

SLI = sim

Preço = R\$ 0.0

Placa Wireless

Fabricante = Killer

Modelo = 1535

Preço = R\$ 29.0

4.2. Resumo da solução número 2

4.2.1. Gráfico de convergência do fitness médio **Figura 4**

Figura 4



Fonte: autor

4.2.2. Especificação do notebook solução número 2

Fitness: 4.352350970769092

Especificação do Notebook

Nome do Modelo = FullRange W1743 IRON V4

Cor

Nome = Azul Fosco

Preço = R\$ 0.0

Sistema Operacional

Descrição = Sem Sistema Operacional

Preço = R\$ 0.0

Office

Tipo = Microsoft Office 2016 Home & Students (Word, Excel, PowerPoint, OneNote)

Preço = R\$ 399.0

Preço padrão

No boleto = R\$ 16996.5

Cartão de crédito = R\$ 18885.0

Desconto = 10.0%

Parcelas = 12

Preço personalizado

No boleto = R\$ 19333.8

Cartão de crédito = R\$ 21482.0

Desconto = 10.0%

Parcelas = 12

Processador

Fabricante = Intel

Família = i7

Modelo = 7700HQ

Memória cache = 6

Frequência Máxima = 3.8

Preço = R\$ 0.0

Memória RAM

Tipo = DDR4

Frequência = 2133.0

Tamanho = 32

Preço = R\$ 465.0

Tela

Polegadas = 15.6

Tecnologia = FullHD 120Hz / 5ms com Tecnologia G-SYNC (Matte)

Preço = R\$ 425.0

1º Memória de armazenamento

Tipo = SSD SATAe M.2

Unidade de medida = GB

Tamanho = 240

Preço = R\$ 0.0

2º Memória de armazenamento

Tipo = HD 5400 RPM

Unidade de medida = TB

Tamanho = 2

Preço = R\$ 566.0

1º SATA_eM2 Memória de armazenamento

Tipo = null

Unidade de medida = null

Tamanho = 0

Preço = R\$ 0.0

2º SATA_eM2 Memória de armazenamento

Tipo = SSD SATAe M.2

Unidade de medida = GB

Tamanho = 480

Preço = R\$ 742.0

Placa de Vídeo

Fabricante = NVIDIA

Modelo = GeForce GTX 1050

Memória dedicada = 2

SLI = false

Preço = R\$ 0.0

Placa Wireless

Fabricante = Intel

Modelo = 3165

Preço = R\$ 0.0

5. Conclusão

O algoritmo desenvolvido tem um funcionamento satisfatório, considerando que pode ajudar um futuro cliente a escolher um notebook de alto desempenho, porém alguns pontos precisam de melhorias, como os seguintes:

- melhorar a função de avaliação de fitness para que se obtenho uma nota de desempenho mais precisa e implementando técnicas de Pareto;
- modificar a forma de pontuação e avaliação de cada componente para que sua nota individual seja armazenada na base de dados ao contrário de ser aplicada em tempo de execução;
- Implementar uma maneira de restrição dos componentes das soluções geradas para evitar a escolha de componentes fora do escopo de um

determinado modelo, pois a técnica de penalização não se mostrou muito eficiente, acarretando a criação de soluções não factíveis;

- Restringir também a inserção de componentes baseando-se na presença de outros para não ocorrer o contexto da solução 2 apresentada, no qual está presente um Office porém o notebook está sem sistema operacional;