

Algoritmo inteligente para otimizar a geração de recomendação alimentar com muitos objetivos

Róger Alves Silva de Araújo, Cícero Garrozi e Rodrigo Gabriel Ferreira Soares

Departamento de Estatística e Informática

UFRPE

Recife, Brasil

Email: rogeralves_90@yahoo.com.br, cicero@gmail.com, rgfsoares@gmail.com

Resumo—Algoritmos evolucionários são utilizados para a obtenção de resultados satisfatórios em uma série de problemas em que algoritmos de busca tradicionais não são capazes de entregar soluções viáveis em um tempo aceitável. Existem exemplos na literatura de trabalhos que utilizaram o potencial destes algoritmos para otimizar a geração de recomendações alimentares para dados critérios nutricionais. Entre as propostas, estão exemplos com o uso de algoritmos mono-objetivo e multiobjetivo, em geral os estudos mostram que resultados nutricionalmente satisfatórios podem ser obtidos mediante o uso de ambas as técnicas, mas problemas relacionados com critérios subjetivos tais como o sabor do prato e o gosto do usuário foram recorrentes nos estudos. No presente trabalho é proposto um algoritmo que modela o problema da geração de recomendação alimentar para muitos objetivos, através do uso de um MaOEA (Many-Objective Evolutionary Algorithm) para a construção de recomendações alimentares, com o acréscimo de conceitos específicos do domínio do problema incorporados em seus operadores, objetivando melhorar os resultados do ponto de vista da combinação lógica dos alimentos para um usuário.

Keywords—Algoritmo; Otimização Multiobjetivo; Recomendação Alimentar; Nutrição

I. INTRODUÇÃO

A necessidade da geração otimizada de recomendações alimentares está relacionada à observação das necessidades apontadas por usuários de sistemas convencionais de acompanhamento nutricional e nutricionistas. Em geral, há usuários de aplicativos móveis que possuem diários de refeições a serem preenchidos manualmente, e nutricionistas que precisam montar um cardápio com opções diversificadas, para pacientes com necessidades distintas, considerando um universo rico de alimentos à disposição. Nestes cenários, a geração de recomendações alimentares automatizadas são muito úteis. Em ambos os casos, muito trabalho manual é necessário: do ponto de vista do paciente, isto leva a um gradual desinteresse em continuar inserindo dados sobre o que ele ingeriu ao longo do dia, o que pode acarretar em uma posterior desistência do programa de acompanhamento; por outro lado, para o profissional nutricionista, o principal problema está no tempo gasto para analisar cada perfil de usuário e montar uma dieta específica que atenda a todas as necessidades nutricionais do paciente. Neste sentido, um sistema de recomendação alimentar que utiliza técnicas de inteligência computacional pode colaborar com ambos. Por um lado, um usuário que recebe uma recomendação pronta e adequada às suas necessidades

possui a facilidade de simplesmente substituir os itens que não gostou, ao invés de ter que realizar a inserção manual, além de se manter motivado por receber recomendações cada vez mais adaptadas ao seu perfil. Por outro lado, os nutricionistas também são beneficiados, pois podem utilizar melhor o seu tempo em tarefas importantes tais como a identificação do perfil nutricional, metas e restrições dos pacientes, já que o trabalho manual de combinar alimentos que sejam adequados a estas metas será realizado pelo sistema.

Em [1] é feito um levantamento das aplicações com maior número de *downloads* na área de acompanhamento nutricional disponíveis nas lojas de aplicativo. Neste levantamento, o autor foi capaz de extrair o fluxo de uso comum deste tipo de aplicativo, o qual apresenta limitações importantes. Dentre os problemas apontados que são recorrentes na literatura estão a ausência de um canal de comunicação entre o usuário e os profissionais de saúde [1] [2] e, principalmente, o método de inserção de dados referentes aos alimentos consumidos, que é manual e desestimulante para o usuário, o que também foi observado em [3] e [4]. Em contrapartida, existem métodos alternativos para registro dos alimentos ingeridos pelos usuários em [4] e [1] que obtiveram resultados satisfatórios mas ainda não há uma solução disponível comercialmente, de modo que as principais aplicações existentes no mercado ainda utilizam o método tradicional de adição manual das informações.

Este trabalho propõe várias modificações a um MaOEA (Many-Objective Evolutionary Algorithm) para resolver este problema de recomendação alimentar. Assim, o nutricionista não precisará definir cardápios, mas utilizará metas nutricionais para o paciente. O uso de um algoritmo evolucionário é adequado também por sua natureza estocástica, ou seja, os resultados obtidos em uma execução não serão necessariamente repetidos nas execuções posteriores. Esta característica é particularmente interessante pois permite oferecer maior variabilidade de opções nas recomendações geradas para o usuário (modelo oposto ao método tradicional que usa fichas de dieta em papel), aumentando a aderência à recomendação já que, com mais opções oferecidas, o usuário poderá substituir itens dos quais não gosta ou não pode ingerir em determinado momento. Além disto, há uma população de soluções não-dominadas que podem ser adotadas como soluções alternativas com qualidade semelhante.

A. Algoritmos evolucionários aplicados à recomendação alimentar

Estão disponíveis na literatura alguns trabalhos com algoritmos evolucionários que foram aplicados no contexto de geração de recomendações alimentares com diferentes abordagens. Em [5] e [6] os autores optaram por usar a otimização mono-objetivo com um AG (algoritmo genético) clássico. Já em [7] e [8] a abordagem multiobjetivo foi aplicada. Ambas as abordagens foram capazes de gerar recomendações satisfatórias do ponto de vista dos objetivos definidos por eles.

Nos trabalhos supracitados, questões subjetivas como o gosto e a lógica do prato foram problemas recorrentes. Para estes critérios, a quantidade de cada nutriente do alimento não é informação suficiente para determinar se dois alimentos fazem sentido em um mesmo prato. A abordagem a estes problemas é geralmente sugerida como trabalhos futuros. Além disso, vale ressaltar que a abordagem para muitos objetivos (MaOEA) apresenta uma oportunidade de exploração neste contexto, visto que os trabalhos anteriores não utilizavam este tipo de algoritmo, que possui implementações já consolidadas tal como o algoritmo NSGA-III [9], para a otimização de recomendações alimentares.

No algoritmo proposto neste trabalho, operadores específicos que visam obter bons resultados do ponto de vista de lógica do prato foram projetados, bem como uma metodologia para o cálculo da aptidão das soluções que adapta certos parâmetros do algoritmo dinamicamente ao longo da evolução de modo que as recomendações apresentadas possuam maior diversidade de opções.

Este artigo está dividido conforme segue. A Seção II apresenta o processo para obtenção do valor energético de uma determinada refeição, usado para a avaliação das recomendações geradas pelo algoritmo proposto. A Seção III aborda o cálculo do valor energético diário, utilizado para a determinação dos objetivos do algoritmo. Na Seção IV são detalhados os componentes que compõem o algoritmo proposto. A Seção V detalha o método de análise dos resultados bem como as métricas utilizadas. Na Seção VI são apresentados e comentados os resultados e, por fim, na Seção VII são apresentadas as conclusões do trabalho e sugestões de trabalhos futuros.

II. O PROCESSO DE OBTENÇÃO DE ENERGIA POR MEIO DOS ALIMENTOS

A alimentação tem dois papéis fundamentais para o organismo humano: prover substâncias essenciais e fornecer energia para manter os processos fisiológicos funcionando. O conjunto de atividades celulares que fica responsável pelo processo de obter energia por meio dos alimentos ingeridos é o Metabolismo. A energia produzida pelo metabolismo é medida em caloria, porém como esta é uma unidade de grandeza muito pequena, é comum que seja utilizado o termo caloria para representar a quilocaloria (Kcal), onde 1 Kcal = 1000 cal.

O metabolismo obtém energia por meio da degradação dos nutrientes dos alimentos que ingerimos. Um simples alimento pode ser composto por diversos nutrientes. Contudo, os que são convertíveis em energia no nosso corpo, os chamados

Nutrientes Energéticos ou Macronutrientes, são os Carboidratos, as Proteínas e os Lipídios. Segundo [10] os valores das taxas de conversão de energia para cada nutriente comumente utilizados equivalem a 4 kcal/g de carboidrato ou proteína e 9 kcal/g de lipídio.

Na Tabela I estão dispostos alimentos que exemplificam uma refeição de café da manhã e seus macronutrientes em gramas. Usando a relação entre a quantidade de calorias por grama de nutriente obtém-se que este prato de café da manhã fictício possui um total de 172,35 Kcal. Por meio deste procedimento é possível então quantificar se uma determinada configuração de prato se aproxima de uma meta nutricional. Em [11] estão estabelecidas diretrizes nacionais para a contribuição da energia proveniente de cada um dos três macronutrientes no valor diário de recomendação energética. Com o uso destas diretrizes, os pratos podem ser avaliados de acordo com os quantitativos de macronutrientes utilizando as mesmas relações utilizadas anteriormente. Por exemplo, se para um determinado indivíduo o valor energético para o café da manhã for de 300 Kcal, das quais 65% devem ser provenientes de carboidratos, 20% de proteínas e 15% de lipídios, então as metas nutricionais em termos de macronutrientes seriam de 48,75 gramas, 15 gramas e 5 gramas, respectivamente.

Tabela I: Exemplo de configuração de prato com seus respectivos macronutrientes

Alimento	Porção	Carboidrato	Proteína	Lipídio
Pão francês	30	17,58	2,4	0,93
Manteiga	10	0,01	0,04	8,24
Café	100	1,5	0,7	0,1

III. CÁLCULO DO VALOR ENERGÉTICO DIÁRIO

Segundo [10], 70% do consumo energético total de um indivíduo corresponde à Taxa Metabólica Basal (TMB) que representa o valor de energia que o corpo humano necessita para manter as suas funções vitais. O consumo energético restante, 30%, está relacionado ao consumo do processo metabólico e à realização de atividades físicas. O método mais usado nas pesquisas para determinar a TMB utiliza equações para o cálculo, tais como em [12] onde é apresentada a equação de Harris e Benedict, que leva em consideração o gênero, o peso corporal, a altura e a idade do paciente:

$$TMB = 447,593 + (9,247P) + (3,098A) - (4,330I) \quad (1)$$

$$TMB = 88,362 + (13,397P) + (4,799A) - (5,677I) \quad (2)$$

As equações (1) e (2) calculam a TMB de acordo com a equação de Harris e Benedict para indivíduos do sexo feminino e masculino, respectivamente, onde P é o peso em quilogramas, A é a altura em centímetros e I corresponde à idade em anos. Outros métodos para o mesmo propósito como as equações da Organização Mundial de Saúde (OMS) estão disponíveis em [10]. Na prática, as equações para o cálculo

Tabela II: Faixas de valores de NAF de acordo com o estilo de vida do paciente

Estilo de vida	NAF
Sedentário a atividades leves	1,40 a 1,69
Ativo a moderadamente ativo	1,70 a 1,99
Intenso a intensamente ativo	2,00 a 2,40

da TMB apresentam resultados semelhantes, portanto neste trabalho utilizou-se a versão de Harris e Benedict.

O custo energético do metabolismo possui baixa participação no valor total das necessidades energéticas e, visto que as fórmulas do TMB fornecem dados aproximados, o metabolismo é desconsiderado no cálculo final da necessidade energética. Assim, uma vez definido o Nível de Atividade Física (NAF), o Valor Energético Diário (VED) pode ser definido conforme Eq. 3.

$$VED = TMB \times NAF \quad (3)$$

O valor do NAF é definido mediante análise das atividades físicas e corresponde à razão entre o TMB e o gasto energético ao longo de um dia. Em [10] é detalhado o processo para a definição da categorização do estilo de vida e definição do NAF de um paciente da Tabela II. Uma vez definidos o NAF e o TMB do paciente, é possível obter o valor energético diário. Como exemplo, uma paciente de 23 anos, 162 cm, pesando 55 kg e com NAF de 1,70, teria através de (1) e (3) o valor energético diário calculado de aproximadamente 2309 Kcal.

IV. ALGORITMO PROPOSTO PARA A OTIMIZAÇÃO

O MaOEA utilizado como base na construção do algoritmo proposto foi o NSGA-III. Analisando o problema da recomendação de alimentos da perspectiva de que cada nutriente (macro ou micro) pode ser um objetivo, nota-se que a quantidade de objetivos pode facilmente ser maior que três. Como exemplo, em [13] existem disponíveis 23 nutrientes para 597 alimentos¹. Uma otimização sobre todos os nutrientes não é esperada na prática, mas é aceitável que se utilize mais do que três deles. Conforme explicado em [9], para este número de objetivos, o NSGA-III apresenta performance reconhecidamente superior em comparação com outros algoritmos multiobjetivo.

A. Definição da solução e sua aptidão

Uma solução candidata para o algoritmo representa uma proposta de prato onde cada gene da solução corresponde ao identificador de um alimento na base de dados, conforme Figura 1 que ilustra a representação de uma refeição contendo quatro alimentos. Uma solução considerada ótima neste contexto é aquela cuja soma de um nutriente específico de todos os alimentos pertencentes a ela seja igual à meta nutricional estabelecida. Para a validação do algoritmo foi definido que os objetivos são equivalentes aos três macronutrientes, isto é,

¹Na prática nem todos os nutrientes são aplicáveis para os 597 alimentos, como é o caso da fibra alimentar para o sal.

13	7	115	64
----	---	-----	----

Figura 1: Exemplo de cromossomo utilizado pelo algoritmo

carboidratos, proteínas e lipídios. O cálculo das metas para cada um destes macronutrientes inicia pela definição do VED conforme (1), (2) e (3), e então obtém-se o valor específico de cada macronutriente usando as proporções mostradas na Seção II. Como o VED representa a recomendação calórica para o dia inteiro e a solução candidata representa uma única refeição, 15% do VED é considerado representando a meta para uma refeição. Três perfis distintos de usuários foram definidos para os testes do algoritmo, cada um com diferentes características físicas e níveis de atividade física.

O cálculo da aptidão é baseado no valor da soma dos nutrientes da solução, convertidos em notas de 0 a 10. Quanto maior a nota de uma solução, maiores as chances dela sobreviver durante a evolução. A estratégia utilizada consiste em monitorar o número da iteração do algoritmo para ajustar o critério de avaliação responsável pela definição da aptidão, utilizando para isso uma gaussiana, e assim reduzir gradualmente a permissividade para as novas gerações, elevando a qualidade dos resultados à medida que a evolução avança. Na configuração da gaussiana, o ponto médio é definido como a meta nutricional desejada e o desvio padrão é definido como a margem de erro aceitável para a meta nutricional, o qual é reajustado ao longo da evolução com valores finais variando entre 10% e 5%. O valor obtido pela função gaussiana é normalizado de modo que a nota máxima é obtida quando o valor da solução for igual à meta nutricional desejada. Assim, a Eq. (4) define a aptidão de uma solução para determinado macronutriente M.

$$fitness_M = \left[\frac{10 \cdot \left(\frac{1}{\sigma \sqrt{2\pi}} \right) \cdot e^{-\frac{(\sum_{i=1}^n N_i - \mu)^2}{2\sigma^2}}}{\frac{1}{\sigma \sqrt{2\pi}}} \right] \times -1 \quad (4)$$

O ajuste da permissividade do algoritmo varia com a alteração do desvio padrão. No início da evolução o seu valor é maior, assim as soluções terão maiores chances de obter boas notas e não serem descartadas pelo algoritmo. À medida que a evolução avança, o desvio padrão vai sendo reduzido, tornando mais rigoroso o critério de seleção, o que leva à melhora da qualidade das soluções sobreviventes. Este processo segue desde a primeira até a última iteração do algoritmo. Conforme pode ser visto na Figura 2, uma solução que obteve nota satisfatória para seguir adiante na primeira geração (Fig. 2a) devido ao relaxamento dos critérios de avaliação, será mal avaliada após algumas gerações e terá poucas chances de chegar ao fim da evolução (Fig. 2b). Até este momento, a solução pode já ter compartilhado bons

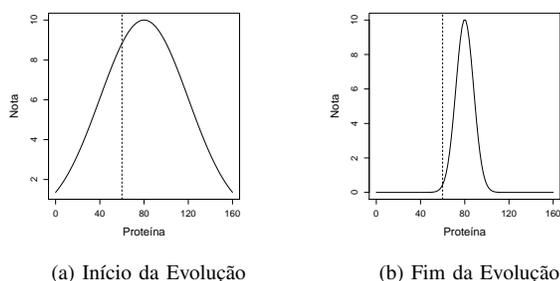


Figura 2: Exemplo de ajuste dinâmico do desvio padrão usado no início e no fim da evolução

genes com outras soluções, fornecendo a sua contribuição positiva para as futuras gerações, privilegiando a diversidade e a fuga de soluções ótimas locais. Na prática, os resultados do algoritmo com esta abordagem de avaliação se mostraram promissores durante os testes, com o aumento do número de recomendações oferecidas.

Uma das limitações apontadas por [5] é o uso de porções de tamanho fixo dos alimentos, acarretando em recomendações limitadas a estas porções. Para contornar este problema, foi adotada a estratégia de permitir duas ou mais porções de um mesmo alimento em uma solução e utilizar recomendações com quantidade variável de alimentos para aumentar o número possível de resultados. Assim, um alimento mais calórico, que não poderia ser indicado em uma recomendação de tamanho fixo por deixar pouco espaço para outros alimentos, passa a ser recomendado para o usuário com o “ônus” de vir acompanhado de menos itens, aumentando assim a possibilidade do usuário aceitar as recomendações fornecidas pelo algoritmo.

Um ponto importante a se destacar refere-se à granularidade das porções. Foram utilizados os dados de [13] como base para o algoritmo realizar a otimização, os quais estão relacionados a uma porção de 100 gramas de cada alimento. A granularidade não é limitada pelo algoritmo proposto, assim basta que novas entradas para outras porções sejam adicionadas na base de dados para que o algoritmo possa utilizá-las. Uma estratégia semelhante a esta é usada pelo aplicativo “Contador de Calorias”, ou no inglês *My Fitness Pal*, que foi citado em [1] como um dos mais populares aplicativos da área de nutrição disponível nas lojas de aplicativos.

Ao decidir por não adotar frações dos alimentos na função objetivo, a representação da solução é simplificada (podendo ser inteira) e os esforços do algoritmo estarão direcionados à otimização dos nutrientes. Por fim, diferentemente dos trabalhos anteriores, o algoritmo proposto é submetido a uma base mais ampla de alimentos e com dados de procedência confiável, obtidos por análise laboratorial em um processo descrito em [13]. A escolha dos dados foi feita com atenção aos trabalhos de [14] e [15], que destacam a atenção relatada por organizações de saúde de alguns países quanto à qualidade e confiabilidade das informações fornecidas por aplicativos

voltados para a área de saúde.

B. Etapas da evolução

1) *Filtragem da população inicial*: A partir da base de alimentos disponíveis para a evolução e de um usuário com suas metas nutricionais estabelecidas, a primeira etapa do algoritmo deverá selecionar os alimentos candidatos a fazer parte da recomendação. Os alimentos que isoladamente ultrapassam 100% da meta nutricional estabelecida em qualquer objetivo não precisam estar presentes em nenhuma etapa da evolução, visto que não fazem parte de uma solução viável. Assim, na criação da população inicial, estes alimentos são desconsiderados, reduzindo o espaço de busca inicial a uma população com indivíduos realmente capazes de gerar um resultado satisfatório.

2) *Cruzamento de soluções com tamanho dinâmico*: O uso de soluções com tamanho dinâmico não foi observado nos algoritmos apresentados nos trabalhos anteriores². Para tal, o operador *SmartNutriOnePointCrossover* foi proposto especificamente para este algoritmo a partir de uma adaptação do cruzamento de um ponto. Ao realizar um cruzamento com dois pais de tamanhos diferentes, o operador deve ter especial atenção ao tamanho das soluções filhas. Neste contexto, os pais contribuem com tamanhos diferentes em cada filho. A Figura 3a mostra a criação de dois filhos “C” e “D” a partir dos pais “A” e “B”, o operador inicialmente realiza a clonagem dos pais e sorteia dois pontos de corte, um para cada filho, de acordo com o tamanho de cada pai. A troca de genes é feita a partir do ponto de corte escolhido até o fim da solução pai que não foi usada para criar inicialmente a solução filha, sobrescrevendo os genes originais até o limite de cruzamento. A principal diferença desta abordagem em relação ao cruzamento de um ponto convencional é que, se os pais forem de tamanhos diferentes, a contribuição do menor pai pode ocorrer através da remoção de alguns genes que foram clonados do pai com maior tamanho, conforme exemplificado na Figura 3a caso o ponto de corte do filho “C” fosse definido a partir do terceiro gene. Dependendo do ponto de corte escolhido, as soluções filhas podem ser mais parecidas com alguma das soluções pai.

3) *Operador híbrido de mutação e reparação*: A mutação é aplicada nos algoritmos evolucionários com o intuito de aumentar a diversidade na população, evitar a convergência prematura e permitir a descoberta de áreas ainda não exploradas utilizando busca local. No presente trabalho, a mutação possui papel crucial para que seja possível realizar uma combinação lógica dos alimentos na refeição. A modelagem da solução utilizando grupos de alimentos por refeição é benéfica ao não misturar alimentos que comumente fazem parte de uma refeição com aqueles de outra refeição. Contudo, como a base de dados utilizada (tabela TACO) é bastante variada, ainda podem ocorrer situações em que itens que fazem parte de um mesmo grupo de alimentos não façam sentido quando combinados em uma mesma configuração de prato.

²Mesmo em outros contextos, algoritmos evolucionários com indivíduos de tamanho variável não são comuns, motivo pelo qual foi necessário adaptar alguns operadores da biblioteca de algoritmos evolucionários utilizada.

Um exemplo seria a recomendação de açúcar. A base de dados utilizada possui três tipos de açúcar: cristal, refinado e mascavo. Não faz sentido em uma mesma configuração de prato ter mais de um tipo de açúcar. Para reduzir estas ocorrências na solução proposta, os alimentos da base de dados utilizada receberam a propriedade “subgrupo” como uma especialização dos grupos alimentares básicos. De fato, em [13] tem-se o grupo de alimentos “Leite e Derivados”, que pode conter os subgrupos: “Bebidas de Leite” e “Laticínios”. A definição dos subgrupos alimentares foi realizada neste trabalho de forma manual, visto que não foi possível estabelecer os subgrupos por meio de técnicas automáticas na base de dados utilizada. O reconhecimento de palavras no nome do alimento foi provado deficiente em [5] mas há a hipótese de que seria possível identificar algum padrão mediante a análise comportamental das refeições reportadas pelos usuários em uma futura implantação do algoritmo proposto em um sistema com usuários reais.

O operador de mutação, denominado no algoritmo de Smart-NutriMutationWithRepair, executa a alteração de uma solução mediante duas abordagens: perturbação genética ponto a ponto e reparação. As duas abordagens são baseadas nos subgrupos, mas possuem objetivos diferentes: a primeira é mais próxima de uma mutação convencional alterando o valor de um gene por outro pertencente ao mesmo subgrupo alimentar e, assim, evitando alimentos que não fazem sentido em uma refeição; a segunda abordagem pode realizar alterações maiores no cromossomo, removendo todas as repetições de subgrupo em uma solução. Diferentemente da abordagem de perturbação genética, a reparação não utiliza uma probabilidade por gene, mas por cromossomo. A Figura 3b ilustra o funcionamento das duas abordagens sobre um cromossomo com quatro genes, cujos alimentos pertencem aos subgrupos “S1”, “S7” e “S10”, mostrando como o algoritmo realiza as alterações nos genes dependendo da estratégia adotada.

4) *Demais parâmetros do algoritmo:* A seleção de sobreviventes ocorre por meio de torneio binário. Neste operador, as soluções candidatas são agrupadas aos pares e as melhores em cada torneio sobreviverão para a próxima geração. Os parâmetros numéricos foram definidos após uma série de testes divididos em ciclos e etapas conforme método a ser abordado na Seção V. Seis parâmetros numéricos foram utilizados para definir uma instância do algoritmo: número de iterações, as probabilidades de cruzamento, mutação e reparação, e as porcentagens de desvio padrão utilizadas no cálculo da aptidão no início e no fim da evolução. Os valores dos parâmetros foram definidos a partir dos trabalhos anteriores. Os valores de desvio padrão inicial e final fornecem certa flexibilidade na montagem das refeições no início das gerações sem impactar na qualidade final dos resultados, fornecendo boas opções disponíveis na curva de Pareto. O número de pontos é particularmente interessante neste problema pois de acordo com [5] a quantidade de opções oferecidas ao usuário aumenta as chances da recomendação ser seguida pelo paciente.

Tabela III: Metas nutricionais para os usuários de teste

Usuário	Perfil	Carboidrato	Proteína	Lipídio
1	Equilibrado	456,45	80,55	59,67
2	Restritivo	313,72	55,36	41,01
3	Permissivo	551,86	97,39	72,14

V. MÉTODO DE ANÁLISE DOS RESULTADOS

A definição da instância do algoritmo que melhor atende ao problema ocorreu por ciclos de teste compostos de 3 etapas. Cada etapa possui o propósito de selecionar as configurações de melhor performance em diferentes cenários. A configuração considerada vencedora é a que apresenta o melhor resultado geral para cada um dos perfis de usuário exibidos na Tabela III ao final das três etapas. Como os valores iniciais de alguns parâmetros numéricos foram definidos empiricamente, ao final do primeiro ciclo de análise realiza-se uma aferição para verificar se a última etapa de testes contém apenas configurações com valores extremos para algum dos parâmetros, o que pode sugerir a possibilidade de melhoria dos resultados com o reajuste dos parâmetros para além dos extremos predefinidos. Neste último caso, um ciclo adicional de testes subsequentes é realizado com todas as 3 etapas para verificar se os resultados serão ainda melhores.

Em cada etapa, são calculados os seguintes valores estatísticos para as métricas de qualidade definidas:

- Média Aritmética: qual valor está equidistante dos extremos obtidos para o indicador ao longo das execuções;
- Mediana: qual valor está no centro dos valores obtidos para o indicador ao longo das execuções. Uma mediana igual a média indica que os valores são distribuídos de forma simétrica;
- Desvio padrão amostral: o quanto os valores obtidos para o indicador oscilam em torno da média. Expressa o nível de estabilidade dos resultados do algoritmo.

Os cálculos estatísticos foram feitos em código Java, utilizando uma classe criada especificamente para o projeto e usando como base as equações consolidadas de cada métrica.

A. Indicadores de desempenho

Comumente, os indicadores de desempenho para algoritmos multiobjetivo são ancorados na relação entre a curva de Pareto do algoritmo com a curva de Pareto ótima. Este conceito se torna por vezes inviável pois a solução ótima pode não ser conhecida para a maioria dos problemas. No caso específico do algoritmo proposto neste trabalho, em que são realizadas adaptações com a finalidade de resolver um tipo específico de problema, é necessário que os indicadores de qualidade sejam definidos de modo a garantir que o algoritmo esteja atingindo as premissas pelas quais ele foi preparado. Partindo deste princípio, foram utilizados cinco indicadores de qualidade para aferir o desempenho das variadas combinações de parâmetros de configuração e dos cenários de teste abordados nas Seções anteriores. Os indicadores foram determinados de modo a avaliar principalmente a diversidade das soluções apresentadas,

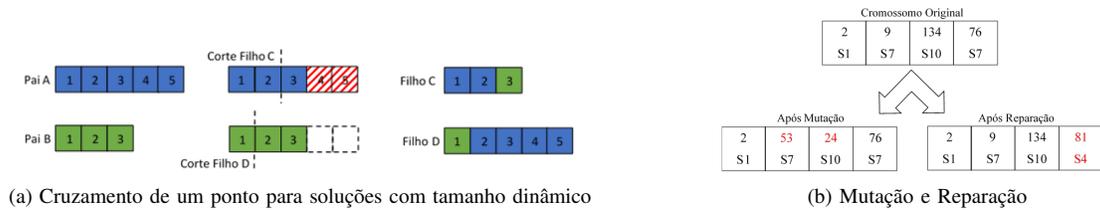


Figura 3: Funcionamento dos operadores personalizados para o algoritmo

a convergência na região de soluções dentro de uma faixa aceitável de desvio dos valores-alvo e a qualidade do ponto de vista da montagem de um prato.

1) *Espalhamento*: Indica o quanto as soluções apresentadas estão afastadas entre si. Em um conjunto de pratos sugeridos em uma refeição, é interessante que existam opções que sejam equivalentemente adequadas do ponto de vista nutricional, mas que apresentem configurações distintas. Assim, aumenta-se a variabilidade de alimentos oferecidos na recomendação e diminui-se a chance do usuário não gostar de nenhuma das opções oferecidas, uma vez que ele terá uma faixa maior de alimentos para escolher. Para efeitos de comparação, ao usar este indicador, uma solução com maior valor de espalhamento é preferível.

2) *Espaçamento*: Usado para verificar a convergência dos resultados, este indicador calcula a distância mínima entre as soluções. Uma vez que o indicador de espalhamento tem relação com os extremos da área coberta pelas soluções, o espaçamento afere se as soluções estão bem distribuídas pela fronteira de Pareto obtida, isto é, se preenchem bem a área demarcada pelos extremos das soluções. A fórmula proposta em [16] foi utilizada para o cálculo do espaçamento. Ao contrário do indicador de espalhamento, uma solução tem melhor espaçamento quando o valor do indicador é menor.

3) *Hipervolume*: O Hipervolume pode ser considerado um indicador híbrido, pois verifica a qualidade da solução em vários aspectos. Seu princípio de funcionamento é baseado como segue: uma vez definido um ponto de referência em uma região oposta à da convergência da evolução, ele calcula a soma do volume compreendido entre este ponto e cada ponto pertencente ao conjunto de Pareto em análise. Valores altos neste operador indicam que as soluções estão bem espaçadas e com bom espalhamento, sendo portando um indicador ratificador dos apresentados anteriormente.

4) *Taxa de acerto*: Este indicador relaciona a quantidade de recomendações pertencentes à curva de Pareto da solução com as soluções que estão em uma região de 10% da meta nutricional estabelecida. É similar ao indicador "Taxa de Erro", o qual verifica a porcentagem de soluções do conjunto de Pareto sob análise que não estão presentes na curva de Pareto ótima. Para o algoritmo proposto, o conceito de curva de Pareto ótima não se aplica, visto que o melhor dos cenários está concentrado em um único ponto teórico do espaço de objetivos que é o ponto (10, 10, 10). O uso da taxa de erro neste contexto não agregaria o valor desejado às métricas de

qualidade, pois a maior parte das soluções teriam 100% de taxa de erro, o que não indica necessariamente que todas as soluções apresentam opções fora do limite de desvio aceitável definido.

5) *Taxa de repetição de subgrupos*: Criado especificamente para este algoritmo, a proposta deste indicador é avaliar a lógica de composição do prato baseado no número de opções oferecidas que apresentam repetição de subgrupos alimentares. Seu cálculo é feito pela divisão do número de soluções que apresentam repetição de subgrupos pelo número total de soluções pertencentes à curva de Pareto.

B. Processo de escolha das configurações

Os ciclos de teste para escolha da instância final do algoritmo foram baseados nos melhores valores para os indicadores de taxa de repetição de subgrupos e taxa de acerto. Conforme relatado anteriormente, as três etapas de análise correspondem a um ciclo de testes, o qual foi repetido algumas vezes até a escolha da instância final do algoritmo. A Figura 4 foi obtida após a análise dos resultados do segundo ciclo de testes e revelou uma relação inversamente proporcional entre as taxas de repetição de subgrupos e de acerto. Portanto, os critérios para escolha da configuração final do algoritmo utilizaram uma terceira rodada de testes e precisaram ser elaborados de modo a não privilegiar demasiadamente um indicador em relação a outro. Os critérios de seleção adotados em cada etapa dos testes neste ciclo estão detalhados a seguir.

1) *Etapa 1*: Em uma primeira etapa, três valores para cada um dos parâmetros numéricos detalhados na Seção IV-B4 foram embaralhados, gerando 729 combinações de configurações, as quais foram executadas para cada um dos três perfis de usuário da Tabela III com dez repetições. Utilizamos o *software* de mineração de dados WEKA para gerar agrupamentos sobre os resultados das execuções de acordo com a similaridade entre eles. Baseado nestes grupos, as configurações que prosseguem para a segunda etapa são escolhidas como segue:

- 1) O agrupamento contendo a faixa com os menores valores de taxa de repetição de subgrupos é escolhido;
- 2) As configurações que obtiveram até 1/3 do valor máximo de taxa de repetição de subgrupos neste agrupamento são ordenadas pela taxa de acerto;
- 3) A metade das configurações com maiores valores de taxa de acerto são selecionadas para a segunda etapa.

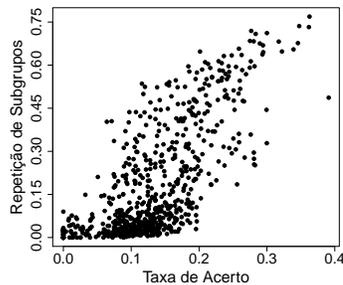


Figura 4: Dispersão dos valores de taxa de acerto e repetição de subgrupo.

2) *Etapas 2 e 3*: A segunda etapa segue os mesmos passos da etapa anterior porém as instâncias do algoritmo são executadas 50 vezes de modo a melhorar a representatividade dos dados estatísticos calculados. Além disso, no último passo apenas as cinco melhores configurações no indicador de taxa de acerto são selecionadas para a terceira etapa. Na etapa final, todos os indicadores de qualidade são considerados, com a configuração final sendo aquela que apresentar o melhor conjunto de resultados em todos eles, para os três perfis de usuário da Tabela III.

VI. RESULTADOS

A execução dos testes no algoritmo conforme o método detalhado na Seção V revelou a existência de uma relação inversamente proporcional entre os indicadores de taxa de acerto e de repetição de subgrupo, conforme Figura 4, onde estão plotados os resultados para as médias de taxa de acerto e repetição de subgrupos referentes às 729 configurações testadas. É possível observar que, quando a taxa de acerto é alta, a taxa de repetição de subgrupos também é alta. Observa-se também que a maneira em que um indicador varia em relação ao outro não é linear, ou seja, é possível reduzir em mais da metade a taxa de repetição de subgrupos reduzindo a taxa de acerto em menos da metade. A partir destes dados pode-se concluir que, uma vez detectados os fatores que causam o aumento da taxa de repetição de subgrupos, é possível reduzir significativamente este indicador com baixa degradação da taxa de acerto.

A análise das configurações do algoritmo pelos resultados obtidos revelou que, de fato, determinados operadores numéricos estão intrinsecamente relacionados com altas taxas de repetição de subgrupos. Após a primeira etapa de testes, as configurações com maiores valores de probabilidade de cruzamento e menores valores de probabilidade de reparação e número de iterações foram eliminadas; a ausência de configurações com estes valores já na segunda etapa de testes revela um padrão para que os melhores resultados sejam obtidos com maior número de iterações e, principalmente, com menor taxa de cruzamento e maior reparação. Os valores finais obtidos que definem a instância do algoritmo proposto estão relacionados abaixo:

- Iterações: 1500
- Probabilidade de cruzamento: 80%
- Probabilidade de mutação: 10%
- Probabilidade de reparação: 90%
- Desvio padrão aceitável no início da evolução: 50%
- Desvio padrão aceitável no fim da evolução: 10%

Para obter resultados mais adequados, é preciso seguir o caminho oposto ao que aumenta o número de soluções aceitáveis. As taxas de acerto e repetição de subgrupos portanto podem ser modeladas em trabalhos futuros como objetivos conflitantes, a fim de buscar uma alternativa mais equilibrada.

Os resultados obtidos nos trabalhos anteriores revelaram que ambas as abordagens mono-objetivo e multiobjetivo eram adequadas para a solução do problema de geração de recomendação alimentar. Portanto, após definida a instância do algoritmo proposto com melhor performance na geração das recomendações alimentares, foi realizada uma comparação entre os resultados obtidos por ele e pelos algoritmos que usam a abordagem mono-objetivo, com o objetivo de demonstrar que esta nova proposta é uma opção viável frente às apresentadas anteriormente. Três abordagens foram desenvolvidas para realizar esta comparação. A primeira abordagem foi inspirada em [6], a segunda foi realizada calculando a aptidão das soluções baseando-se na diferença em gramas entre a meta nutricional e o que foi recomendado pelo algoritmo para um nutriente específico, e a terceira foi semelhante à segunda porém considerando a combinação de todos os nutrientes definidos como objetivo do algoritmo proposto em um só objetivo. Os dois últimos algoritmos citados foram desenvolvidos especificamente para os testes.

Na Figura 5, “A1” representa o algoritmo proposto e as três abordagens mono-objetivo supracitadas estão representadas respectivamente por “A2”, “A3” e “A4”. Comparando as abordagens mono-objetivo com o algoritmo proposto, o destaque ocorreu na diferença do indicador de taxa de repetição de subgrupos. Neste indicador, os algoritmos mono-objetivo obtiveram resultados consideravelmente inferiores. Vale ressaltar, porém, que nenhum dos algoritmos mono-objetivo utilizados contavam com um operador de reparação para auxiliar na construção de soluções melhores do ponto de vista lógico. Contudo, o impacto que a técnica de reparação das soluções candidatas alcançou foi significativo na melhoria da qualidade do prato oferecido ao usuário, o que confirma a hipótese inicial que incluir conhecimentos do problema em etapas da evolução pode levar a resultados melhores em comparação ao uso de um algoritmo clássico.

VII. CONCLUSÕES E TRABALHOS FUTUROS

As recomendações montadas pelo algoritmo proposto neste trabalho se mostraram capazes de fornecer uma quantidade considerável de recomendações alimentares diferentes e adequadas nutricionalmente às especificações definidas. Os testes mostraram que, dado um grupo alimentar personalizado e um conjunto de metas nutricionais, o algoritmo é capaz de gerar recomendações que atendam às condições fornecidas. Evidenciou-se que, com a variação das metas nutricionais por

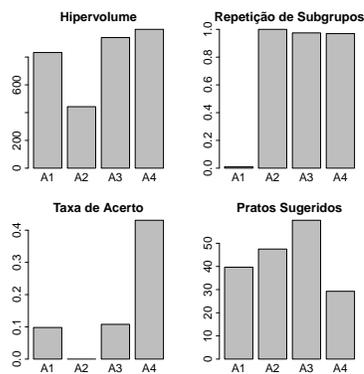


Figura 5: Comparativo entre métricas obtidas por meio das abordagens mono e multiobjetivo

meio do uso de perfis diferentes de usuário, o algoritmo se manteve capaz de gerar recomendações adequadas.

A modelagem do algoritmo foi realizada de modo a possibilitar que o grupo e os subgrupos alimentares utilizados durante a evolução sejam dados como entradas no sistema. Isto possibilita que o nutricionista possa definir livremente os grupos de alimentos adequados a cada perfil de usuário, cabendo ao algoritmo realizar a combinação dos itens de forma adequada. Os resultados adequados também mostraram que a adição de conhecimentos específicos do problema nas diversas etapas da evolução foi uma decisão acertada. Com o uso de soluções de tamanho variável, amplia-se o número de opções de pratos possíveis e torna-se possível a participação de alimentos que, por possuírem grandes quantidades de algum nutriente específico, não seriam oferecidos em uma abordagem convencional de tamanho fixo. Vale destacar também a participação ativa do operador de reparação na melhoria da lógica dos pratos montados conforme observado na Figura 5.

As principais contribuições do algoritmo proposto em relação aos demais algoritmos testados ocorreram na lógica do prato, que é em parte atribuído ao operador de reparação, e na dispersão dos resultados no espaço objetivo. A distribuição dos resultados obtidos pelas abordagens mono-objetivo é geralmente irregular, com os pontos condensados em pequenas regiões próximas dos extremos de cada objetivo. Com o uso da abordagem multiobjetivo, há soluções com maior diversidade de itens, aumentando a chance de alguma recomendação agradar ao usuário conforme relatado por [5].

Por fim, é importante considerar que o algoritmo proposto deve ser empregado como parte de um sistema maior de recomendação alimentar que inclui o desenvolvimento de aplicações clientes (web e mobile) e outros algoritmos acessórios, para tornar possível o seu uso comercial. Trabalhos futuros também abordarão incluir preferências do usuário no processo de escolha de alimentos, de modo a refinar os resultados, além de processos de mineração de dados que auxiliem na detecção de combinações lógicas de alimentos baseado em padrões, como por exemplo na localidade, perfil social ou período do ano, dentre outros.

REFERÊNCIAS

- [1] R. Z. Franco, R. Fallaize, J. A. Lovegrove, and F. Hwang, "Popular nutrition-related mobile apps: A feature assessment," *JMIR mHealth and uHealth*, vol. 4, no. 3, p. e85, Aug. 2016. [Online]. Available: <https://doi.org/10.2196/mhealth.5846>
- [2] M. Jensen, C. Quesada-López, G. Zúñiga, A. Chinnock, and M. Jenkins, "Design, development and evaluation of a mobile application for goal setting and self-monitoring of dietary behaviors," *Journal of the Academy of Nutrition and Dietetics*, vol. 115, no. 9, p. A67, Sep. 2015. [Online]. Available: <https://doi.org/10.1016/j.jand.2015.06.240>
- [3] S. Gabrielli, M. Dianti, R. Maimone, M. Betta, L. Filippi, M. Ghezzi, and S. Forti, "Design of a mobile app for nutrition education (TreC-LifeStyle) and formative evaluation with families of overweight children," *JMIR mHealth and uHealth*, vol. 5, no. 4, p. e48, Apr. 2017. [Online]. Available: <https://doi.org/10.2196/mhealth.7080>
- [4] N. Hongu, B. T. Pope, P. Bilgiç, B. J. Orr, A. Suzuki, A. S. Kim, N. C. Merchant, and D. J. Roe, "Usability of a smartphone food picture app for assisting 24-hour dietary recall: a pilot study," *Nutrition Research and Practice*, vol. 9, no. 2, p. 207, 2015. [Online]. Available: <https://doi.org/10.4162/nrp.2015.9.2.207>
- [5] C.-H. Chen, M. Karvela, M. Sohbaty, T. Shinawatra, and C. Toumazou, "PERSON—personalized expert recommendation system for optimized nutrition," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 12, no. 1, pp. 151–160, Feb. 2018. [Online]. Available: <https://doi.org/10.1109/tbcas.2017.2760504>
- [6] T. Isokawa and N. Matsui, "Performances in GA-based menu production for hospital meals," in *2015 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, May 2015. [Online]. Available: <https://doi.org/10.1109/cec.2015.7257195>
- [7] M. I. V. Segismundo and B. E. V. Comendador, "Prenatal nutrition diet generator utilizing modified genetic algorithm for smartphone," *Journal of Automation and Control Engineering*, pp. 56–60, 2015. [Online]. Available: <https://doi.org/10.12720/joace.3.1.56-60>
- [8] Z. Pei and Z. Liu, "Nutritional diet decision using multi-objective difference evolutionary algorithm," in *2009 International Conference on Computational Intelligence and Natural Computing*. IEEE, Jun. 2009. [Online]. Available: <https://doi.org/10.1109/cinc.2009.175>
- [9] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints," *IEEE Transactions on Evolutionary Computation*, vol. 18, no. 4, pp. 577–601, Aug. 2014. [Online]. Available: <https://doi.org/10.1109/tevc.2013.2281535>
- [10] L. Cuppari, *Nutricao Clinica no Adulto*, 4th ed. MANOLE (SAUDE -TECNICO) - GRUPO MANOLE, aug 2019, iSBN: 8520433294.
- [11] E. A. Melo, P. C. Jaime, and C. A. Monteiro, *Guia alimentar para a população brasileira*, 2nd ed. Ministério da Saúde, 2014, iSBN: 978-85-334-2176-9.
- [12] A. M. Roza and H. M. Shizgal, "The harris benedict equation reevaluated: resting energy requirements and the body cell mass," *The American Journal of Clinical Nutrition*, vol. 40, no. 1, pp. 168–182, Jul. 1984. [Online]. Available: <https://doi.org/10.1093/ajcn/40.1.168>
- [13] *TACO: Tabela Brasileira de Composição de alimentos*. NEPA-UNICAMP, 2011. [Online]. Available: http://www.cfn.org.br/wp-content/uploads/2017/03/taco_4_edicao_ampliada_e_revisada.pdf
- [14] M. I. F. Aranda, "Nutrition apps for pregnant women in primary care, key aspects for use," *Nutrición clínica y dietética hospitalaria*, no. 35, p. 75–79, 2015. [Online]. Available: <http://doi.org/10.12873/352faranda>
- [15] J. Shuren, B. Patel, and S. Gottlieb, "FDA regulation of mobile medical apps," *JAMA*, vol. 320, no. 4, p. 337, Jul. 2018. [Online]. Available: <https://doi.org/10.1001/jama.2018.8832>
- [16] D. A. Van Veldhuizen, "Multiobjective evolutionary algorithms: Classifications, analyses, and new innovations," Ph.D. dissertation, Air Force Institute of Technology, Wright Patterson AFB, OH, USA, 1999, aA19928483.