JORDI ALVES REINSMA

Orientador: Marco Antonio Moreira de Carvalho

MÉTODO EXATO PARA RESOLUÇÃO DO PROBLEMA DE MINIMIZAÇÃO DE BLOCOS DE UNS CONSECUTIVOS

Ouro Preto

Julho de 2018

Universidade Federal de Ouro Preto
Instituto de Ciências Exatas
Bacharelado em Ciência da Computação

MÉTODO EXATO PARA RESOLUÇÃO DO PROBLEMA DE MINIMIZAÇÃO DE BLOCOS DE UNS CONSECUTIVOS

Monografia apresentada ao Curso de Bacharelado em Ciência da Computação da Universidade Federal de Ouro Preto como requisito parcial para a obtenção do grau de Bacharel em Ciência da Computação.

JORDI ALVES REINSMA

Ouro Preto
Julho de 2018

UNIVERSIDADE FEDERAL DE OURO PRETO

FOLHA DE APROVAÇÃO

Método Exato para Resolução do Problema de Minimização de Blocos de Uns Consecutivos

JORDI ALVES REINSMA

Monografia defendida e aprovada pela banca examinadora constituída por:

Dr. Marco Antonio Moreira de Carvalho – Orientador
Universidade Federal de Ouro Preto

Dr. Marcone Jamilson Freitas Souza
Universidade Federal de Ouro Preto

Dr. Túlio Ângelo Machado Toffolo
Universidade Federal de Ouro Preto

Ouro Preto, Julho de 2018

Resumo

O problema de Minimização de Blocos de Uns Consecutivos (ou COB, do inglês *Consecutive Ones Block Minimization Problem*) é um problema $\mathcal{NP}$-Difícil que consiste em determinar a permutação das colunas de uma matriz binária que minimize as descontinuidades de blocos de elementos não nulos em cada uma de suas linhas. Este problema é estudado de maneira independente sob diferentes nomes em diferentes áreas do conhecimento, como pesquisa operacional, arqueologia, microbiologia, sistemas de informação e matemática. Este estudo reporta a aplicação da redução do COB a uma versão do conhecido Problema do Caixeiro Viajante (*Traveling Salesman Problem*, TSP), baseado em resultados parciais de diferentes autores presentes na literatura. Após a redução, são propostos dois métodos – um exato e um heurístico – para a resolução do TSP.

Abstract

The Consecutive Ones Block (COB) Minimization is an $\mathcal{NP}$-Hard problem which consists in finding a permutation of the columns of a binary matrix such that the discontinuities of its blocks of nonzero elements are minimized in each row. This problem is independently studied under different names in different knowledge areas, such as operations research, archeaology, microbiology, information systems and mathematics. This study reports the reduction of COB to a version of the Traveling Salesman Problem (TSP), based on partial results from different authors found in the literature. After the reduction, two methods are proposed – an exact method and a heuristic – for solving the TSP.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Algoritmos

Capítulo 1

Introdução

Dentre os vários problemas já abordados na literatura de otimização inteira, há uma série destes, encontrados principalmente em contextos industriais, que são abstraídos e modelados em matrizes binárias. O Problema de Determinação de Leiaute de Matrizes de Portas (*Gate Matrix Layout Problem*, GMLP), o Problema de Minimização de Espalhamento de Ordens de Compra (*Minimization of Order Spread Problem*, MORP) e o Problema de Minimização de Pilhas Abertas (*Minimization of Open Stacks Problem*, MOSP) são exemplos desta modelagem em contextos industriais. O GMLP está relacionado ao projeto de circuitos eletrônicos com dimensões otimizadas, o MORP está relacionado com o tempo de atendimento a pedidos de compra e o MOSP está relacionado com a utilização otimizada de estoque intermediário. Da mesma maneira, problemas combinatórios com a mesma modelagem podem ser encontrados na matemática. Cita-se o Problema de Minimização de Banda (*Bandwidth Minimization Problem*), Problema de Largura de Corte (*Cutwidth*) e Problema de Aumento de Uns Consecutivos (*Consecutive Ones Augmentation Problem*).

Neste trabalho é discutido o Problema de Minimização de Blocos de Uns Consecutivos (*Consecutive Ones Blocks*, COB), em que, a partir de uma matriz binária, deve-se obter o menor número de blocos de uns nas linhas da matriz, utilizando-se de permutações das colunas. Define-se *blocos de uns* como uma sequência de uns consecutivos em uma linha de uma matriz, na qual a presença de zeros interrompe tal sequência. Para a resolução deste problema, é proposta a aplicação de um método exato a partir de uma modelagem clássica da literatura para o COB. Desta maneira, é possível gerar os resultados ótimos para instâncias da literatura pela primeira vez. Adicionalmente, propõe-se a aplicação de um método heurístico para a solução do COB a partir da mesma modelagem clássica. Por fim, é discutida a relação deste problema com os demais citados anteriormente, assim como a aplicação direta do COB na indústria e em outros contextos.

1.1 Motivação

O COB é um problema da classe $\mathcal{NP}$-Difícil (Garey e Johnson, 1979), possuindo, portanto, uma grande complexidade teórica. A dificuldade de se trabalhar com problemas com esta característica é uma motivação para encontrar métodos para sua melhor resolução, acrescentando conhecimento à literatura. Adicionalmente, este problema possui similaridades de resolução com outros problemas de matrizes binárias, como será comentado na revisão da literatura.

O Problema de Minimização de Descontinuidades (*Minimization of Discontinuities Problem*, MDP) é exemplo de uma aplicação do COB no contexto industrial. Neste problema, o objetivo é a redução do número de interrupções na fabricação sequencial de produtos. Estas interrupções podem ocasionar ociosidade da linha de produção, ou mesmo diferenças nas características físicas dos produtos. Em ambos os casos, as interrupções influenciam os custos relacionados à produção. Assim, o estudo de métodos de solução do COB tornam-o interessante economicamente para empresas de pequeno, médio e grande porte.

O problema de reduzir a quantidade de "espaços" entre "elementos" também ocorre na arqueologia (Kendall, 1969) – com a rotulação sítios, microbiologia (Alizadeh et al., 1995) – com a reconstrução de moléculas de DNA a partir de clones dos fragmentos da molécula, compressão de estruturas de dados (Johnson et al., 2004) e na organização de arquivos (Kou, 1977). A multidisciplinaridade do problema também é um fator motivacional para sua abordagem.

1.2 Objetivos

Em linhas gerais, o objetivo deste trabalho é, a partir de algoritmos eficientes e consistentes que possam ser utilizados no contexto do problema abordado, produzir soluções ótimas ou próximas da solução ótima, inicialmente considerando problemas específicos, mas com uma possibilidade de generalização. São objetivos específicos:

1. Realizar pesquisa para geração de embasamento teórico e revisão bibliográfica sobre o tema tratado;

2. Realizar pesquisa sobre a modelagem do Problema de Minimização de Blocos Consecutivos de Uns como o Problema do Caixeiro Viajante.

3. Implementar a modelagem e aplicar pela primeira vez um método exato na resolução do problema original;

4. Avaliar o método utilizado considerando dados reais e também com problemas teste publicamente disponíveis;

5. Ampliar o conjunto de problemas testes com soluções ótimas já conhecidas através da execução de testes sistemáticos usando formulações já desenvolvidas;

6. Utilizar método heurístico para avaliação de instâncias patológicas, ou seja, instâncias em que o método exato falhe em produzir solução em tempo hábil;

7. Produzir trabalhos para serem publicados em periódicos e eventos científicos.

1.3 Organização do Trabalho

O restante deste trabalho está organizado como segue. O Capítulo 2 apresenta a literatura relevante ao problema estudado. Em seguida é abordada a fundamentação teórica do COB, no Capítulo 3. O Capítulo 4 detalha a modelagem utilizada, assim como as ferramentas usadas para apoiar a resolução do problema. Por fim, é comentado o plano de atividades futuras no Capítulo 5 e as conclusões deste trabalho, no Capítulo 6.

Capítulo 2

Revisão da Literatura

O Problema de Minimização de Blocos de Uns Consecutivos é estudado há mais de 40 anos, sendo considerado em diferentes contextos e utilizando-se métodos diversos para a resolução do problema. Neste capítulo, são consideradas as pesquisas mais relevantes encontradas na literatura sobre o COB e outros problemas relacionados. São mantidas as denominações originais utilizadas em cada trabalho. Também é discutida a relação entre o COB e estes outros problemas permutacionais que também envolvem matrizes binárias.

A primeira pesquisa analisada relativa ao COB é a de Dyson e Gregory (1974). No trabalho, foi realizado um estudo da produção de chapas de vidro de uma fábrica inglesa, visando solucionar o Problema de Corte de Estoque (*Cutting-Stock Problem*) em combinação com o COB, este último sendo abstraído como Problema de Alocação de Padrões (*Pattern Allocation Problem*). Primeiro, o Problema de Corte foi resolvido via programação linear e, a partir dos padrões de corte obtidos, a fase de sequenciamento dos padrões foi modelada como o Problema do Caixeiro Viajante (*Traveling Salesman Problem*, TSP). Esta modelagem foi descartada, visto que o TSP não possuía resolução ótima realizável em tempo hábil na época da pesquisa. Foi então proposto um algoritmo *branch-and-bound* modificado para gerar os padrões de corte e o sequenciamento. A função objetivo utilizada penaliza as descontinuidades e prioriza sequências de padrões consideradas difíceis de ordenar por análise empírica. Também, são priorizados os padrões que produzem peças já inseridas na sequência, a fim de finalizá-las rapidamente, minimizando a ocorrência de interrupções.

Kou (1977) e Garey e Johnson (1979) apresentaram provas formais de que o COB pertence à classe *NP-Difícil*. Anos depois, Haddadi (2002), estudando o COB sob o nome Problema de Minimização de Blocos Consecutivos (*Consecutive Blocks Minimization Problem*, CBMP), demonstra que o problema permanece sendo *NP-Difícil* mesmo em matrizes binárias restringidas a possuir exatamente dois elementos não nulos por linha.

O COB é abordado novamente em conjunto com o Problema de Corte de Estoque na pesquisa de Madsen (1979), no contexto de uma pequena empresa produtora de chapas de vidro. Após a obtenção dos padrões de corte via resolução do Problema da Mochila (*Knapsack*

Problem), é utilizada a representação de matriz binária, quadrada e simétrica, para indicar se padrões possuem peças em comum ou não. Em seguida, é aplicada uma heurística originalmente aplicada ao problema de Largura de Banda de Matrizes Simétricas (*Matrix Bandwidth*) para a redução da distância de produção e, consequentemente, a minimização das interrupções. Em uma pesquisa seguinte sob o mesmo contexto, Madsen (1988) transforma a instância de COB advinda dos padrões de corte em uma instância de TSP, obtendo então uma solução utilizando uma heurística sub-ótima do TSP.

Haddadi e Layouni (2008) discutem o problema de Minimização de Blocos Consecutivos (*Consecutive Blocks Minimization*, CBM), uma aplicação do COB para a redução da complexidade de resolução de sistemas de equações lineares. Neste trabalho, é demonstrado como construir uma solução com fator de aproximação de 1,5 do valor da solução ótima para o problema. A transformação se baseia em reduções entre problemas que não alteram a solução do problema original. A instância de CBM, uma matriz binária de n linhas e m colunas é transformada no problema de Minimização de Blocos Circulares (*Circular Blocks Minimization*, CIR), que relaxa a interpretação da matriz no sentido de que as colunas são consideradas em ordem circular. Consequentemente, os blocos são também circulares. Em seguida, é feita a transformação do CIR para uma instância simétrica do TSP. A matriz de distâncias criada, baseada na distância entre as colunas do problema original, é equivalente à *distância de Hamming*, o que caracteriza o *Hamming Traveling Salesman Problem* (HTSP). A partir desses dados, é utilizado o algoritmo de Christofides (1976), cujo fator de aproximação é de 1,5 para encontrar uma solução para o TSP na instância transformada.

Novamente sob o contexto de redução das dimensões de sistemas lineares, Haddadi et al. (2015) propoem duas buscas locais para o CBM com complexidade $O(n^2)$, sendo n a quantidade de linhas da matriz binária do problema. A primeira busca local é a troca da posição de duas colunas (método *2-swap*), e a segunda é a remoção de uma coluna de sua posição para inserção em outra posição, deslocando as demais colunas (método *shift*). Ambas as buscas locais são implementadas com um método de avaliação do impacto de modificar soluções com complexidade de tempo $O(m)$, em que m representa o número de colunas da matriz binária. Esse método denominado Δ-avaliação, constitui uma contribuição importante, uma vez que a avaliação completo de uma matriz binária exige complexidade $\Theta(mn)$. Para comparar o desempenho das heurísticas, foram utilizadas 5 instâncias reais de uma companhia ferroviária alemã, além de 45 instâncias geradas aleatoriamente com diferentes características. As instâncias artificiais, gentilmente cedidas pelo autor, serão utilizadas para avaliar o desempenho do algoritmo produzido neste trabalho, comparando-o com os resultados obtidos por demais autores.

Nascimento e Carvalho (2017) introduzem uma representação em grafos do COB, estudado sob o contexto do MDP. A modelagem considera as peças a serem produzidas como sendo os vértices, possuindo adjacência entre si se as peças estão presentes em um mesmo padrão. Adicionalmente, o peso de cada aresta indica o número de vezes em que cada par de peças é produzido a partir de um mesmo padrão. Utiliza-se um algoritmo de busca em largura modificado para gerar uma solução inicial de maneira construtiva, priorizando vértices com menor número de adjacências. A partir da solução inicial, a metaheurística *Iterated Local Search* é empregada com diferentes buscas locais e um método de perturbação para obtenção de refinamento da solução. Foram consideradas as instâncias artificiais produzidas por Haddadi et al. (2015) e seus resultados para a avaliação do desempenho do método proposto. A metaheurística encontrou melhores soluções para as instâncias densas, ou seja, matrizes com grande quantidade de elementos não nulos. Para instâncias pouco densas, o método não conseguiu se aproximar dos resultados do estudo de referência.

Há na literatura diversos problemas combinatórios envolvendo matrizes binárias. A extensa pesquisa de Dom (2009) apresenta esta diversidade, mostrando suas correlações. É válido destacar que há também problemas idênticos sendo estudados sob nomes e contextos diferentes, com pesquisas não possuindo conhecimento das partes. De maneira similar Linhares e Yanasse (2002) relacionaram diferentes problemas modelados em matrizes binárias. Neste trabalho, estabeleceu-se a equivalência entre alguns problemas, como MOSP e GMLP, entretanto, foi demonstrado que o MDP não é equivalente aos problemas MOSP, GMLP, MORP e ao Problema de Minimização de Trocas de Ferramentas (*Minimization of Tool Switches Problem*, MTSP), este último definido como o problema de minimizar a quantidade de operações de troca de ferramentas de uma máquina flexível em uma linha de produção.

O trabalho de Chakhlevitch et al. (2013) contempla o Problema do Aumento de Uns Consecutivos (*Consecutive Ones Augmentation Problem*, C1AP) que trata minimizar a quantidade de zeros entre uns em matrizes binárias. Foram introduzidos três algoritmos: um algoritmo exato de busca extensiva e duas heurísticas construtivas. A primeira heurística, denominada *GRIN*, descreve uma inserção gulosa (*GReedy INsertion*), onde é inserida uma nova linha à solução por vez, analisando a posição da inserção que menos aumenta a função objetivo. A segunda heurística, *MAZE*, constrói a solução pela primeira e última linha da matriz, fazendo a inserção de demais linhas entre as mesmas visando maximizar o número de zeros nas extremidades verticais da matriz (*MAximizing the number of end-ZEros*). Para a análise dos três métodos, foram criadas 50 instâncias artificiais aleatórias para cada diferente tamanho de matriz, variando entre 4×4 e 10×1023. Também foram produzidas 5 instâncias de 25 colunas e 20000 linhas. O algoritmo exato possui complexidade de tempo $O(mn(n!)^2)$ para m linhas e n colunas da matriz. Dada sua complexidade, este algoritmo foi executado apenas nas instâncias com 8 ou menos colunas, determinando seus resultados ótimos. Por fim, as heurísticas *GRIN* e *MAZE* foram comparadas entre si em todas as instâncias, nas quais o método *MAZE* obteve resultados majoritariamente melhores em relação ao *GRIN*, produzindo soluções ótimas globais em todas as instâncias com resultados ótimos conhecidos. É importante destacar que o C1AP é o mesmo problema estudado pelo MORP, que busca a redução do tempo entre a produção da primeira peça até a requisição da última peça de um mesmo tipo para todas as peças em uma linha de produção. Estes são exemplos de problemas idênticos porém com estudos não interligados até o momento deste trabalho.

Capítulo 3

Fundamentação Teórica

Este capítulo aborda a descrição detalhada do COB, criando uma rígida base de conceitos necessária para produção dos métodos computacionais deste trabalho. Também são descritos dois problemas relacionados, que, por possuírem forte relação ao COB, terão suas características específicas levadas em consideração para a elaboração do algoritmo.

3.1 O Problema de Minimização de Blocos de Uns Consecutivos

Formalmente, o COB considera uma matriz binária $A_{m \times n}$, i.e., $A = \{a_{ij}\}$, com $a_{ij} \in \{0, 1\}$, $i = \{1, 2, ..., m\}$ e $j = \{1, 2, ..., n\}$. Em uma dada linha de A, uma sequência contígua de qualquer número de elementos não nulos (inclusive um único) é denominada *bloco de uns consecutivos*. A existência de um elemento nulo entre dois blocos de uns consecutivos caracteriza uma *descontinuidade*. A existência de um único bloco de uns consecutivos em uma linha de A implica na inexistência de descontinuidades. De forma análoga, a existência de dois ou mais blocos de uns consecutivos em uma mesma linha implica na existência de uma ou mais descontinuidades, respectivamente.

Uma matriz $A_{6 \times 6}$, exemplo de uma instância do COB, é apresentada a seguir. Nesta matriz, a linha 1 contém dois blocos de uns, separados por uma descontinuidade presente na coluna b. Na linha 4 há três blocos – o primeiro na coluna a, outro em d e o último em f. No total, a matriz A possui 11 blocos de uns e 5 descontinuidades, sendo estas últimas destacadas em negrito.

Uma solução para o COB é representada por uma permutação π das colunas da matriz de entrada A. Desta forma, é possível determinar a matriz permutação A^{π}, que possui as colunas de A na ordem estabelecida por π. São apresentados dois exemplos de solução para a matriz A. A permutação da matriz A^{π_1} é representada pela sequência de colunas $\pi_1 = [b, a, c, e, d, f]$ e a permutação da matriz A^{π_2} pela sequência $\pi_2 = [d, f, c, a, b, e]$. Nota-se que em ambas as

$$A = \begin{array}{c c} & \begin{array}{cccccc} a & b & c & d & e & f \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} & \begin{pmatrix} 1 & \mathbf{0} & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & \mathbf{0} & \mathbf{0} & 1 & \mathbf{0} & 1 \\ 0 & 1 & \mathbf{0} & \mathbf{0} & 1 & 0 \\ 0 & 0 & 1 & \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \end{array}$$

matrizes produzidas o número de blocos consecutivos de uns e de descontinuidades se alterou, havendo três descontinuidades na matriz A^{π_1} e duas na matriz A^{π_2}. Considerando isso, o objetivo do COB é a minimização do número de blocos. O tamanho do espaço de soluções do COB é limitado por $O(n!)$. Sendo n o número de colunas da matriz binária, o número total de soluções existentes é de $\frac{n!}{2}$, levando em conta que permutações em ordem inversa produzem solução idêntica à sua contraparte.

$$A^{\pi_1} = \begin{array}{c c} & \begin{array}{cccccc} b & a & c & e & d & f \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} & \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & \mathbf{0} & \mathbf{0} & 1 & 1 \\ 1 & \mathbf{0} & \mathbf{0} & 1 & 0 & 0 \\ 0 & 0 & 1 & \mathbf{0} & \mathbf{0} & 1 \end{pmatrix} \end{array} \qquad A^{\pi_2} = \begin{array}{c c} & \begin{array}{cccccc} d & f & c & a & b & e \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} & \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & 1 \\ 1 & 1 & \mathbf{0} & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

3.2 O Problema de Minimização de Blocos de Uns Circulares

Sendo um problema intimamente relacionado ao COB, o *Circ*BM também tem como objetivo minimizar o número de blocos de uns consecutivos em uma matriz binária. Entretanto, este problema tem a peculiaridade de que os elementos da última coluna da matriz são considerados adjacentes aos elementos na primeira coluna de suas respectivas linhas. Em outras palavras, as colunas são consideradas dispostas circularmente.

Ao analisar a mesma matriz A utilizada no exemplo da Seção 3.1 porém interpretando-a ao *Circ*BM, seu número de blocos consecutivos torna-se 10, considerando que os uns da primeira e última coluna da quarta linha de A são adjacentes entre si, exemplificado abaixo.

$$A = \begin{array}{c c} & \begin{array}{cccccc} a & b & c & d & e & f \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} & \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

3.3 O Problema do Caixeiro Viajante

O Problema do Caixeiro Viajante (*Traveling Salesman Problem*, TSP) é caracterizado por um conjunto de n cidades e um conjunto D de distâncias entre todas as cidades. Em sua versão de otimização, o problema é $\mathcal{NP}$-Difícil e consiste em determinar uma rota de visitação das cidades com menor distância total e de forma a visitar cada cidade exatamente uma vez, retornando à cidade inicial no fim da rota. Em outras palavras, uma solução do TSP pode ser entendida como uma permutação circular do conjunto de cidades de menor custo. Os tipos de distância consideradas caracterizam versões diferentes do TSP, e podem ser rodoviárias, euclidianas, *Manhattan* ou outras, dependendo de sua aplicação.

Abaixo observa-se um exemplo de instância do TSP, uma matriz $D_{5\times 5}$ de distâncias entre 5 cidades. O elemento d_{ij} representa a distância entre a cidade i e a cidade j. Nota-se que, neste exemplo, a distância entre uma cidade e outra é idêntica à distância de volta, i.e. $d_{ij} = d_{ji}$ para toda combinação de i e j. Isto caracteriza uma instância *simétrica* de TSP.

$$D = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \end{array} \\ \begin{pmatrix} 0 & 30 & 60 & 90 & 50 \\ 30 & 0 & 60 & 70 & 40 \\ 60 & 60 & 0 & 40 & 80 \\ 90 & 70 & 40 & 0 & 50 \\ 50 & 40 & 80 & 50 & 0 \end{pmatrix} \end{array}$$

Uma solução para o TSP é representada por uma permutação S de n cidades e seu custo total de distância. Para a matriz de distâncias, o custo total é encontrado pela equação abaixo.

$$d_{s[n]s[1]} + \sum_{i=1}^{n-1} d_{s[i]s[i+1]}$$

Para a matriz D a sequência $[1, 2, 5, 4, 3]$, por exemplo, é avaliada como $d_{31} + d_{12} + d_{25} + d_{54} + d_{43}$, resultando em uma rota de custo de 220 unidades.

Capítulo 4

Desenvolvimento

Uma forma comum de contornar as dificuldades de solucionar um problema computacional é transformando-o em um outro problema para o qual haja mais recursos disponíveis para sua abordagem. Esta transformação deve permitir que a solução obtida possa ser transformada de volta ao problema original, com a garantia de que a qualidade da solução adquirida se mantenha. No texto a seguir, a redutibilidade de um problema A a um problema B com preservação de aproximação em tempo polinomial (*polynomial-time approximation scheme*)é denotada por $A \leq_{PTAS} B$.

4.1 Redução do COB ao HTSP

É apresentada a redução do COB ao HTSP, conforme introduzida e provada parcial ou totalmente por diferentes autores (Kou, 1977; Alizadeh et al., 1995; Greenberg e Istrail, 1995; Haddadi e Layouni, 2008). Esta redução é realizada em tempo determinístico polinomial e de modo que o COB e o HTSP podem ser considerados problemas equivalentes. Desta maneira, um método de solução para o HTSP pode ser aplicado de maneira equivalente para solução do COB.

A redução é realizada em duas etapas. Na primeira delas, o COB é reduzido ao Problema de Minimização de Blocos Circulares (*Circular Block Minimization*, *Circ*BM). Na segunda etapa, o *Circ*BM é reduzido ao HTSP.

4.1.1 Redução ao Problema de Minimização de Blocos Circulares

Conforme Haddadi e Layouni (2008), para reduzir o COB ao *Circ*BM, modifica-se a matriz binária A com a inserção de uma nova coluna artificial preenchida apenas com elementos nulos, i.e., $B = [0 \cup A]$. Desta forma, a matriz $A_{m \times n}$ dá origem a uma matriz $B_{m \times n+1}$, contendo uma coluna nula. Esta transformação é simples, assim como a demonstração de sua equivalência.

Lema 1. *COB* $\leq_{PTAS}$ *CircBM.*

Demonstração. *Sem perda de generalidade, utilizaremos como exemplo a matriz A. A permutação* $\pi_1 = [a, f, c, d, e, b]$ *é uma solução ótima de A para o CircBM, denotada por* $A^{\pi 1}$*. Para o COB, esta permutação não é ótima, justificada pela existência de soluções melhores, como mostrada na Seção 3.1.*

$$A = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} a & b & c & d & e & f \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \end{array} \qquad A^{\pi 1} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{cccccc} a & f & c & d & e & b \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Adicionando-se uma coluna de elementos nulos a A e resolvendo o CircBM nesta nova matriz B, encontramos o resultado ótimo $\pi_2 = [a, f, c, 0, d, e, b]$*. Pela característica circular do problema, a permutação* $\pi_3 = [0, d, e, b, a, f, c]$ *é idêntica à* π_2*. Devido à descontinuidade em todas as linhas causada pela coluna nula, não ocorrem blocos consecutivos circulares. Isto torna* π_3 *uma permutação ótima para o COB também. Adicionalmente, a remoção da coluna nula da matriz resultante não altera esta solução, considerando que não há blocos de uns à esquerda desta coluna. Assim, mostramos que uma instância COB consegue ser transformada em uma CircBM com a preservação da otimalidade de sua solução e, sem perda de generalidade, comprovamos também tal propriedade.*

$$B^{\pi 2} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{ccccccc} a & f & c & \boldsymbol{0} & d & e & b \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & \mathbf{0} & 0 & 0 & 0 \\ 1 & 0 & 0 & \mathbf{0} & 0 & 0 & 1 \\ 0 & 0 & 0 & \mathbf{0} & 1 & 1 & 0 \\ 1 & 1 & 0 & \mathbf{0} & 1 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{0} & 0 & 1 & 1 \\ 0 & 1 & 1 & \mathbf{0} & 0 & 0 & 0 \end{pmatrix} \end{array} \qquad B^{\pi 3} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{ccccccc} \boldsymbol{0} & d & e & b & a & f & c \end{array} \\ \begin{pmatrix} \mathbf{0} & 0 & 0 & 0 & 1 & 0 & 1 \\ \mathbf{0} & 0 & 0 & 1 & 1 & 0 & 0 \\ \mathbf{0} & 1 & 1 & 0 & 0 & 0 & 0 \\ \mathbf{0} & 1 & 0 & 0 & 1 & 1 & 0 \\ \mathbf{0} & 0 & 1 & 1 & 0 & 0 & 0 \\ \mathbf{0} & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

4.1.2 Redução ao Problema do Caixeiro Viajante

Para redução do *Circ*BM ao TSP, utiliza-se a distância de *Hamming*, que caracteriza o HTSP. Esta distância originalmente é utilizada para determinar o grau de similaridade entre duas cadeias de caracteres de mesmo tamanho. A distância é calculada em função do número de caracteres em que duas cadeias se diferenciam entre si e é naturalmente simétrica, i.e., a distância de uma cadeia para outra é idêntica à distância inversa. Adicionalmente, a distância de *Hamming* obedece à desigualdade triangular (Robinson, 2003, pp. 255-257).

A prova formal da preservação da qualidade da solução proveniente desta transformação é descrita detalhadamente por Haddadi e Layouni (2008).

Lema 2. *CircBM* $\leq_{PTAS}$ *TSP.*

Demonstração. *Nesta redução, considera-se cada uma das $n+1$ colunas da matriz B como uma cidade do HTSP. As distâncias D_h entre as cidades são dadas pelas distâncias de Hamming entre as colunas de B, i.e., considera-se cada coluna como uma cadeia binária de caracteres e a distância é dada pelo número de não correspondências entre os elementos de cada uma de suas linhas. Intuitivamente, quanto menor a distância de Hamming de duas colunas de B, maior o desejo de produzir uma permutação em que estas colunas aparecem contiguamente na solução.*

A matriz de distâncias D_h pode ser construída, a partir da matriz B, pelo seguinte cálculo: $D_h = B^T(M^1 - B) + (M^1 - B)^T B$, conforme Haddadi e Layouni (2008), na qual M^1 representa uma matriz $m \times n+1$ constituída inteiramente de 1s, e ^{T} indica a função de transposição de matriz. Este algoritmo possui complexidade de tempo $O(n^3)$, por se tratar de uma sequência de multiplicação de matrizes. Contudo, é possível implementar o cálculo de distâncias de *Hamming* em complexidade de tempo $O(n^2 m)$, ao interpretar cada coluna da matriz de entrada como sendo uma cadeia de m bits e aplicando a operação XOR para cada combinação de pares de cadeias de bits. O número de elementos não-nulos na cadeia de bits resultante equivale à distância *Hamming* entre tais colunas.

Nos casos em que o número de linhas é menor do que o número de colunas, a forma proposta para cálculo das distâncias de *Hamming* possui menor complexidade assintótica. Outra vantagem computacional é a utilização de operações bit a bit, mais rápidas na prática quando comparadas à multiplicações de números inteiros.

A solução do HTSP para a instância $(n+1, D_h)$ é equivalente à solução do *Circ*BM para a instância B, conforme provado por (Kou, 1977; Alizadeh et al., 1995; Greenberg e Istrail, 1995). Haddadi e Layouni (2008) comprovam também que o custo da rota da solução HTSP é igual ao dobro do número de blocos circulares consecutivos de uns para o *Circ*BM. Por transitividade, esta solução também é equivalente à solução do COB para a instância A, como queríamos demonstrar.

Teorema. *COB $\leq_{PTAS}$ TSP.*

A matriz D_h ilustra a redução do *Circ*BM ao HTSP, considerando a matriz B.

$$B = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array} \begin{array}{c} \begin{array}{ccccccc} \mathit{0} & a & b & c & d & e & f \end{array} \\ \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \end{array} \qquad D_h = \begin{array}{c} \\ \mathit{0} \\ a \\ b \\ c \\ d \\ e \\ f \end{array} \begin{array}{c} \begin{array}{ccccccc} \mathit{0} & a & b & c & d & e & f \end{array} \\ \begin{pmatrix} 0 & 3 & 2 & 2 & 2 & 2 & 2 \\ 3 & 0 & 3 & 3 & 3 & 5 & 3 \\ 2 & 3 & 0 & 4 & 4 & 2 & 4 \\ 2 & 3 & 4 & 0 & 4 & 4 & 2 \\ 2 & 3 & 4 & 0 & 2 & 2 & \\ 2 & 5 & 2 & 4 & 2 & 0 & 4 \\ 2 & 3 & 4 & 2 & 2 & 4 & 0 \end{pmatrix} \end{array}$$

Dada a correspondência entre COB e HTSP, é possível utilizar qualquer algoritmo de resolução do TSP simétrico para produzir uma solução para o COB. A literatura de tais algoritmos é vasta, possuindo métodos de aproximação (Christofides, 1976), heurísticas subótimas de bom desempenho (Lin e Kernighan, 1973) e métodos exatos (Held e Karp, 1962).

4.2 O Resolvedor Concorde

O método considerado estado-da-arte por mais de 15 anos entre os métodos exatos para solução do TSP simétrico é o resolvedor Concorde (Applegate et al., 2006). Este programa é escrito na linguagem ANSI C, e é disponível gratuitamente para fins acadêmicos, necessitando de uma licença para seu uso comercial. Sua biblioteca é constituída por mais de 700 funções, permitindo sua integração a códigos de algoritmos específicos de problemas relacionados ao TSP.

O resolvedor Concorde produz soluções ótimas para o TSP utilizando algoritmos de programação linear inteira e programação dinâmica, busca de limitantes inferiores para o valor solução e diversas heurísticas auxiliares para obtenção de limitantes superiores. Dentre os métodos empregados, destacam-se a produção de planos de cortes e relaxações, a triangulação de *Delaunay* e algoritmos para determinar árvores geradoras mínimas. Para a fase de produção de planos de cortes, é necessária a utilização de um resolvedor de programação linear auxiliar. O Concorde possui suporte para dois resolvedores: o *QSopt* e o CPLEX.

Desta forma, utiliza-se neste trabalho a modelagem do COB como o HTSP e resolve-se este último problema por meio do resolvedor exato Concorde e, posteriormente, obtém-se o número de descontinuidades. A escolha deste programa é adequada devido à propriedade das distância de *Hamming* serem simétricas e obedecerem a desigualdade triangular. Embora alguns trabalhos da literatura tenham utilizado anteriormente esta modelagem em determinados contextos, as soluções foram geradas por meio de heurísticas.

A solução ótima obtida pelo Concorde para o HTSP consiste em uma permutação π^* de $n+1$ elementos. A solução ótima para o COB é obtida a partir da permutação das colunas da matriz B em razão da ordem estabelecida por π^* e por fim pela remoção da coluna de elementos nulos. A solução ótima da instância de HTSP para a matriz D_h da Seção anterior é $\pi = [\mathit{0}, d, e, b, a, f, c]$. Esta sequência é então aplicada à matriz com a coluna extra B^{π^*}. A redução para o COB é enfim feita com a remoção desta coluna, produzindo a matriz A^{π^*}. As soluções ótimas *Circ*BM e COB são apresentadas abaixo.

$$
B^{\pi *} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{array}{c}
\begin{array}{ccccccc} \boldsymbol{0} & d & e & b & a & f & c \end{array} \\
\begin{pmatrix}
\mathbf{0} & 0 & 0 & 0 & 1 & 0 & 1 \\
\mathbf{0} & 0 & 0 & 1 & 1 & 0 & 0 \\
\mathbf{0} & 1 & 1 & 0 & 0 & 0 & 0 \\
\mathbf{0} & 1 & 0 & 0 & 1 & 1 & 0 \\
\mathbf{0} & 0 & 1 & 1 & 0 & 0 & 0 \\
\mathbf{0} & 0 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}
\end{array}
\qquad
A^{\pi^*} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{array}
\begin{array}{c}
\begin{array}{cccccc} d & e & b & a & f & c \end{array} \\
\begin{pmatrix}
0 & 0 & 0 & 1 & \mathbf{0} & 1 \\
0 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 \\
1 & \mathbf{0} & \mathbf{0} & 1 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}
\end{array}
$$

4.3 A heurística LKH

LKH é uma implementação aprimorada da heurística *Lin-Kernighan* (Lin e Kernighan, 1973), devida a Helsgaun (2000). Assim como o Concorde, seu código é escrito em ANSI C e de uso gratuito para fins acadêmicos e para uso não-comercial. O método é extremamente eficiente, apesar de ser um algoritmo heurístico, detendo recordes de melhor solução encontrada para instâncias de dimensões muito grandes cujos valores ótimos das soluções são desconhecidos. Entre estas instâncias cita-se a *world TSP*, que contém 1.904.711 cidades. Outro recorde é devido à solução da instância contendo 85.900 cidades, para a qual o método obteve a solução ótima, determinada anteriormente pelo resolvedor Concorde.

A heurística Lin-Kernighan consiste na utilização de uma busca local da solução do TSP chamada *k-opt variável*, em que k arestas pertencentes à atual solução são removidas e substituídas por outras k arestas, considerando um grafo de entrada do problema. A cada passo do algoritmo, considera-se a execução de um *k+1-opt* através de regras heurísticas de seleção de vértices, a fim de melhorar a solução. Este processo continua, reiniciando o valor de k ao fim de cada passo para um valor padrão, até que o critério de parada do algoritmo seja satisfeito. Em uma implementação ingênua, uma busca *k-opt* possui complexidade de tempo $O(n^k)$, o que torna o algoritmo computacionalmente inviável para grandes valores de n ou de k. Na etapa de análise da execução do *k+1-opt*, são selecionadas heuristicamente cidades-chave para a participação da busca, no intuito de acelerar o tempo computacional.

Helsgaun aprimora as regras heurísticas originais para a seleção de cidades-chave para a busca, e emprega estratégias de buscas maiores e mais complexas. Por exemplo, menciona-se o particionamento em subproblemas, a implementação eficiente da busca *k-opt variável*, a utilização de outras buscas locais auxiliares e a utilização de análise de sensibilidade para direcionar e restringir a busca. Desta forma, produziu-se um método híbrido robusto.

Como já explicado em detalhe na Seção 4.2, utiliza-se o LKH para determinar soluções heurísticas π^h para instâncias COB transformadas em instâncias HTSP, possibilitando a análise de tempo e qualidade de soluções em relação ao resolvedor Concorde.

Capítulo 5

Plano de Atividades Futuras

Visando a continuação e conclusão deste projeto na disciplina Monografia II, propõe-se um planejamento das atividades subsequentes a este trabalho. A concentração do próximo trabalho se dará por implementar os métodos de solução do COB e posterior realização de experimentos computacionais para as instâncias presentes na literatura, analisando o potencial do método via comparação com o atual estado da arte para o problema tratado. É também planejado a produção de instâncias próprias para o problema, para aprofundar a análise sobre instâncias com características mais abrangentes.

A Tabela 5.1 apresenta o cronograma das atividades restantes, a serem realizadas no segundo semestre de 2018.

Tabela 5.1: Cronograma das atividades futuras.

Etapas do Projeto	Meses			
	01	02	03	04
1. Redação da versão final da Monografia I	✓			
2. Implementação dos métodos de solução	✓	✓		
3. Execução de testes computacionais		✓		
4. Produção de novas instâncias		✓		
5. Execução de testes computacionais		✓	✓	
6. Descrição dos experimentos			✓	
7. Análise dos experimentos			✓	✓
8. Conclusão do trabalho				✓

Capítulo 6

Conclusões

Este trabalho introduz um método exato e um heurístico para o Problema de Minimização Blocos de Uns Consecutivos (ou COB, do inglês *Consecutive Ones Block Minimization*), um problema $\mathcal{NP}$-Difícil estudado em diferentes áreas do conhecimento, de maneira independente. Os métodos consistem inicialmente na redução do COB ao Problema de Minimização de Blocos Circulares, posteriormente reduzido ao Problema do Caixeiro Viajante com distâncias *Hamming* (ou *Hamming distance Traveling Salesman Problem*, HTSP). Após realizar as reduções, aplica-se o resolvedor exato Concorde e o método heurístico LKH à solução do HTSP, cujas soluções são finalmente transformadas nas soluções ótimas e aproximadas do COB, respectivamente.

Os trabalhos futuros serão concentrados em gerar novos conjuntos de instâncias e na execução de experimentos computacionais para averiguar o potencial dos métodos discutidos.

Referências Bibliográficas

Alizadeh, F.; Karp, R. M.; Newberg, L. A. e Weisser, D. K. (1995). Physical mapping of chromosomes: A combinatorial problem in molecular biology. *Algorithmica*, 13(1):52–76.

Applegate, D.; Bixby, R.; Chvatal, V. e Cook, W. (2006). Concorde TSP solver.

Chakhlevitch, K.; Glass, C. A. e Shakhlevich, N. V. (2013). Minimising the number of gap-zeros in binary matrices. *European Journal of Operational Research*, 229(1):48–58.

Christofides, N. (1976). Worst-case analysis of a new heuristic for the travelling salesman problem. Technical report, Carnegie-Mellon Univ Pittsburgh Pa Management Sciences Research Group.

Dom, M. (2009). *Recognition, Generation, and Application of Binary Matrices with the Consecutive Ones Property*. Cuvillier.

Dyson, R. e Gregory, A. (1974). The cutting stock problem in the flat glass industry. *Journal of the Operational Research Society*, 25(1):41–53.

Garey, M. R. e Johnson, D. S. (1979). *Computers and intractability*. WH Freeman New York.

Greenberg, D. S. e Istrail, S. (1995). Physical mapping by STS hybridization: Algorithmic strategies and the challenge of software evaluation. *Journal of Computational Biology*, 2(2):219–273.

Haddadi, S. (2002). A note on the NP–hardness of the consecutive block minimization problem. *International Transactions in Operational Research*, 9(6):775–777.

Haddadi, S.; Chenche, S.; Cheraitia, M. e Guessoum, F. (2015). Polynomial-time local-improvement algorithm for consecutive block minimization. *Information Processing Letters*, 115(6):612–617.

Haddadi, S. e Layouni, Z. (2008). Consecutive block minimization is 1.5-approximable. *Information Processing Letters*, 108(3):132–135.

Held, M. e Karp, R. M. (1962). A dynamic programming approach to sequencing problems. *Journal of the Society for Industrial and Applied Mathematics*, 10(1):196–210.

Helsgaun, K. (2000). An effective implementation of the lin–kernighan traveling salesman heuristic. *European Journal of Operational Research*, 126(1):106–130.

Johnson, D.; Krishnan, S.; Chhugani, J.; Kumar, S. e Venkatasubramanian, S. (2004). Compressing large boolean matrices using reordering techniques. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, pp. 13–23. VLDB Endowment.

Kendall, D. (1969). Incidence matrices, interval graphs and seriation in archeology. *Pacific Journal of mathematics*, 28(3):565–570.

Kou, L. T. (1977). Polynomial complete consecutive information retrieval problems. *SIAM Journal on Computing*, 6(1):67–75.

Lin, S. e Kernighan, B. W. (1973). An effective heuristic algorithm for the traveling-salesman problem. *Operations research*, 21(2):498–516.

Linhares, A. e Yanasse, H. H. (2002). Connections between cutting-pattern sequencing, VLSI design, and flexible machines. *Computers & Operations Research*, 29(12):1759–1772.

Madsen, O. B. (1979). Glass cutting in a small firm. *Mathematical Programming*, 17(1):85–90.

Madsen, O. B. (1988). An application of travelling-salesman routines to solve pattern-allocation problems in the glass industry. *Journal of the Operational Research Society*, 39(3):249–256.

Nascimento, L. H. L. e Carvalho, M. A. M. (2017). Uma heurística aplicada à uniformidade das características físicas de produtos. In *Anais do XLIX Simpósio Brasileiro de Pesquisa Operacional*.

Robinson, D. J. (2003). *An introduction to abstract algebra*. Walter de Gruyter.