

O Uso de Técnicas de Aprendizado de Máquina Combinadas com um Modelo Matemático de Otimização para a Solução do Problema Combinatório de Localização de Instalações

Samuel Willian Alves Wu, Arnaldo Rabello de Aguiar Vallim Filho

Faculdade de Computação e Informática – Universidade Presbiteriana Mackenzie

São Paulo – SP - Brasil

{samuelwuw@hotmail.com, aavallim@mackenzie.br}

Abstract. *This project seeks to solve the warehouse location problem in the logistic area, using machine learning technics, more specifically artificial neural networks and self-organizing maps (SOM) together with mathematical modelling to optimize the results. The problem in question treats of demand points the needs to be attended by warehouses, those that must be generated by the algorithm having as a parameter the demand points location given by latitude and longitude, and the group of points demand level, given by population, for example. Once the warehouses are defined, a mathematical model realizes the task of select the bests using the fixed cost of each installation and the transportation variable cost, both also calculated by the algorithm.*

Resumo. *Este projeto busca resolver o problema de localização de instalações na área de logística utilizando-se de técnicas de aprendizado de máquina, mais especificamente redes neurais artificiais e mapas auto organizáveis (SOM) em conjunto de modelagem matemática para a otimização dos resultados. O problema em questão trata de pontos de demanda que devem ser atendidos por instalações, estas que devem ser geradas pelo algoritmo tendo como parâmetro a localização dos pontos de demanda dado por latitude e longitude, e o nível de demanda do grupo de pontos, dado por população, por exemplo. Uma vez definidas as instalações, um modelo matemático realiza a tarefa de selecionar as melhores utilizando os cálculos de custo fixo de cada instalação e de custo variável de transporte, ambas também calculadas pelo algoritmo.*

1. Introdução

Este trabalho é uma continuação direta de um estudo referente ao problema combinatório de localização de instalações (Wu e Vallim Filho, 2020). O problema é clássico da área de logística, e segundo Ballou (2006), representa a mais importante decisão estratégica de uma operação desta área, além de ser um problema combinatório conhecido na área de Mineração de Dados (Silva et al., 2016).

O objetivo é desenvolver um algoritmo que utilize abstrações de técnicas de redes neurais artificiais (RNA) capazes de realizar a análise de agrupamento dos pontos atendidos pelos centroides (instalações) e a seguir, tendo como parâmetro custos fixos e variáveis, combinar essa solução com um modelo matemático de otimização baseado em programação binária gerando uma saída gráfica no modelo de gráfico de estrelas. Sendo assim o algoritmo é testado com novas bases de dados de diferentes quantidades de pontos de demanda a serem atendidos, a fim de testar a eficiência do algoritmo.

O estudo de localizações tem sua importância comprovada pelo grande número de aplicações que são encontradas em muitas áreas de atividade. Brandeau e Chiu (1989) ilustram essa versatilidade em sua pesquisa que envolve uma grande relação de aplicações nas áreas privada e pública. É importante citar alguns na área privada: depósitos e centros de produção; terminais de transporte e centros logísticos de distribuição de carga; plantas industriais; redes de comunicações; centrais elétricas; centrais de táxi. Já na área pública, tem-se: centrais de veículos de emergência (ambulâncias, bombeiros, etc.); centros de serviços públicos em geral, tais como: centros de saúde, bancos de sangue, centrais de tratamento de lixo, centrais de tratamento de água; localização de bairros residenciais e até mesmo, instalações de defesa.

1.1. Problema de Pesquisa

Contextualizando o problema de localização no contexto de operações logísticas, a quais este trabalho tem foco, pode-se dizer resumidamente que o problema contempla um conjunto de pontos de demanda distribuídos em uma área geográfica, que devem ser atendidos por instalações logísticas. A demanda de cada ponto, dada em unidades de peso, volume ou outra unidade, devem ser satisfeitas por uma instalação, estas tendo de atender um conjunto de pontos

Seguem alguns exemplos de instalações logísticas:

- . Unidades de Produção (fábricas);
- . Centros de Armazenagem de Matéria Prima;
- . Centros de Armazenagem de Produto Acabado;
- . Centros de Distribuição;
- . Operadores Logísticos;
- . Outros.

Uma vez definido o conceito das instalações, podemos tratar da questão básica de localização de instalações que tem como principais objetivos:

- .Quantidade de instalações;
- .Localização das instalações;
- .Alocação de pontos de demanda de cada instalação;

O problema da localização é comumente resolvido minimizando um índice que multiplica a distância percorrida pela quantidade de carga movimentada ou então, minimizando o custo logística da operação.

Observe que este é um problema de otimização combinatória levando em conta a grande quantidade de alternativas disponíveis viáveis. Por ser um problema NP-Difícil, as técnicas de solução geral utilizam metaheurísticas, pois algoritmos exatos não são capazes de resolver problemas de grande porte como este, consequentemente aumentando o tempo computacional de forma exponencial.

Entretanto, apesar de ser comum o uso de metaheurísticas na resolução de problemas de otimização combinatória, a pesquisa tem como foco o uso de redes neurais artificiais (RNA) e modelagem matemática de programação binária para a resolução deste problema. Uma vez que se obtêm a solução ótima do problema, há a tarefa de se desenvolver uma apresentação gráfica do problema em forma de gráfico estrela. Desta forma, podemos sintetizar a questão central deste estudo, sendo ela: “como se comporta a solução de um problema da localização quando abordada por técnicas de rede neural artificial (RNA) combinadas com um modelo matemático de otimização, quais as dificuldades e facilidades de implementação dessas técnicas para solucionar esse tipo de problema, e como pode ser implementada uma apresentação gráfica da solução obtida?”.

2. Referencial Teórico

O objetivo central desta sessão é apresentar os trabalhos científicos que possuem relação com o tema do presente projeto, de forma a auxiliar no entendimento do problema e também dar suporte de conhecimento das estratégias de solução que tratam a questão de localização de instalações.

Como apontado por Vallim Filho (2004), o problema da localização tem sido comumente resolvido por modelo exatos de Programação Matemática (PM) e por metaheurísticas. Problemas de porte pequeno e médio podem utilizar a PM para serem resolvidos, entretanto problemas de grande porte, que são os mais frequentes na prática, são resolvidos por metaheurísticas, já que conforme dito acima, o se configuram como NP-Difícil.

Neto et al. (2010) definiu o termo “metaheurística” derivando sua composição de duas palavras gregas: o prefixo “meta” tendo o significado de “após”, e “heurística” remetendo ao significado de um conhecimento circunstancial, ou seja, compreendendo a etapa de uma camada superior do processo de descoberta de uma solução para um determinado problema. Metaheurísticas são técnicas que abrangem classes mais amplas de problemas. Em outras palavras, é possível utilizar uma metaheurística que pode ser usada para resolver problemas de localização de instalações, para resolver problemas de roteirização de veículos, por exemplo. É importante notar que uma metaheurística na verdade não é um método pronto para solucionar um problema, tais métodos que de acordo com Pirlot (1996), se constituem, na verdade, em “princípios gerais de busca organizados em estratégias gerais“, não podendo assim, ser considerados como algoritmos, mas sim realmente como métodos ou então como meta-algoritmos. Sendo assim, uma metaheurística pode ser considerada como uma estratégia geral de solução de uma classe de problemas. Uma estrutura geral de solução, que uma vez implementada computacionalmente se torna um framework.

São algoritmos aproximados que incorporam mecanismos para evitar confinamento em mínimos ou máximos locais. Conhecimentos específicos do problema

podem ser utilizados na forma de heurística para auxiliar no processo de busca (por exemplo, na busca de um possível bom vizinho de um determinado ponto). Em síntese, pode-se dizer que metaheurísticas são mecanismos de alto nível para explorar espaços de busca, cada uma usando um determinado tipo de estratégia (Becceneri, 2012).

Colorni et al. (1997) define as técnicas as classificando entre a fronteira da pesquisa operacional e a inteligência artificial.

Dentre as diversas metaheurísticas disponíveis para estudo, Laporte (2009) define as principais categorias, sendo elas:

- . metaheurísticas focadas em busca local, como por exemplo: simulated annealing, deterministic annealing e tabu search
- . metaheurísticas de busca em população, tais como: adaptive memory procedures, genetic search
- . metaheurísticas de busca em população, tais como: adaptive memory procedures,
- . mecanismos de aprendizado, tais como: RNA e colônia de formigas.

Levando para o contexto da área de logística, Vallim Filho (2004) demonstrou o uso da metaheurística Simulated Annealing (SA) na tarefa de determinar a localização ótima de Centros de Distribuição (CD). Tendo como parâmetro os custos de localização e de transporte, em uma rede com dois estágios de transferência para as CD's e entrega nos destinos finais, uma função objetiva é criada a fim de resolver o problema encontrando uma solução ótima.

Pirlot (1996) explica o funcionamento da técnica SA em seu artigo, juntamente com o funcionamento das metaheurísticas busca tabu e algoritmos genéticos, apresentando a descrição e as aplicações destas técnicas. Colorni et al. (1997) também apresentam tutoriais com a descrição dos principais aspectos das metaheurísticas derivadas da natureza disponibilizando uma revisão e uma aplicação de SA. Uma aplicação da SA em um problema eminentemente prático foi feita por D'Amico et al. (2002), abordando o problema de redistribuição ou redesenho dos limites do comando policial utilizando a metaheurística, consequentemente definindo as localizações das delegacias. Comparando com o problema deste trabalho, os distritos definidos por D'Amico se comparam ao conjunto de pontos de demanda, e as delegacias se comparam às instalações.

O problema de localização também é aplicado ao setor militar, sendo estudado por Bell (2003) em seu trabalho que teve como objetivo fornecer uma solução para a melhor forma de reposição de inventário de munições da Força Aérea dos EUA, ação necessária para futuros conflitos e para uma maior variedade de cenários de demanda. Uma análise estatística da aplicação do AS no problema da p-mediana é fornecida por Chiyoshi e Galvão (2000), no qual a localização da instalação é definida pela mediana. A técnica é posta à prova com um conjunto de problemas, cada um executado com no máximo cem fluxos diferentes de números aleatórios, condição necessária para a aplicação da metaheurística. Outro artigo interessante relacionado ao tema é o de Abdinour-Helm (2001), que trata do problema p-hub, em que os hubs são as instalações, deve-se definir essas localizações para diferentes valores de p. O artigo mostra uma solução pela metaheurística AS.

Em outra publicação Silva e Cunha (2017) descrevem uma metaheurística de busca tabu (TS) para o problema de cobertura máxima do p-hub. O objetivo é determinar a melhor localização para p hubs e a atribuição de cada um dos pontos de demanda a um único hub, de modo que a demanda total entre pares de nós, dentro de uma determinada distância de cobertura, seja maximizada.

Em um dos poucos trabalhos baseados em RNA, Matsuda e Yoshimoto(2003) apresentam uma abordagem de solução por meio de RNA. O algoritmo parte de uma solução inicial e depois encontra a solução ótima por meio de uma rede baseada em mapas auto-organizáveis (self organizing maps – SOM).

Os mapas auto-organizáveis foram foco de Teuvo Kohonen em seu estudo na década de 1980 e são baseados sobretudo no princípio da formação de mapas de unidades cerebrais (neurônios), como apontado por Silva et al (2016). De forma autoorganizada, estabelecem uma ordenação espacial que permite representar informação. A arquitetura é representada por duas camadas: uma composta por neurônios sensoriais sendo responsável pela entrada de dados e pelos estímulos enviados para a rede neural.

Um dos primeiros estudos envolvendo o problema de localização de instalações foi feito por Hakimi (1964), explorando o problema no contexto das redes telefônicas da época e de estações policiais. O seu estudo procurou por meio de modelos matemáticos encontrar a melhor localização para se instalar centros de transferência do Sistema de telefonia da época, tendo em vista que nesse tipo de rede de comunicações, se tinha a presença de um centro de transferências a qual todas as mensagens precisavam passar antes de se chegar no seu destino. Sendo assim, o objetivo do problema é encontrar a localização exata para os centros de transferência, de modo que o gasto com fios de comunicação seja minimizado. Hakimi também apontou a similaridade do problema com a localização exata de centros de polícia e hospitais, demonstrando o número de aplicações que o problema pode ser colocado.

Zhang e Rushton (2008) mais recentemente exploraram o problema de localização de instalações em um contexto competitivo de sistemas de serviço com múltiplos centroides a serem definidos, maximizando o espaço útil dos usuários levando em conta os limitantes de tempo de espera e o orçamento dos proprietários do sistema. O objetivo geral do trabalho foi desenvolver um modelo de suporte que auxiliasse nas decisões de proprietários de redes de atendimento com múltiplas sedes. Tendo-se um número de filiais já instaladas, o estudo buscou formas e modelos matemáticos capazes de escolher o melhor local para se instalar uma nova filial para uma das firmas.

Uma análise dos métodos localização de instalações no contexto de sistemas de serviço de saúde foi realizada por Rahman e Smith (1999), a fim de verificar a eficácia do método no planejamento desses sistemas em nações em desenvolvimento. É apontado que embora estes métodos matemáticos sejam muito sofisticados para seu uso em nações em desenvolvimento, diversos estudos demonstram que tais métodos tem uma real utilidade no processo de decisão de localizações. O Objetivo central do trabalho se divide em: examinar o papel de modelos de localização de instalações; considerar sua relevância em problemas de desenvolvimento em nações em desenvolvimento afinal, como apontado no trabalho, muitas decisões neste cenário são

tomadas localmente por oficiais do governo e/ou líderes elegidos, fazendo com que a maioria dessas decisões esteja muito distante do que pode ser considerado ótimo.

Os modelos que utilizam metaheurísticas também podem ser utilizados no problema de sinais de trânsito. A otimização do tempo de sinais de trânsito em redes de estradas foi tema de um estudo realizado por Oda et al. (1997), utilizando a técnica de *Simulated annealing* para encontrar soluções ótimas. Este é um problema combinatório de otimização, ou seja, neste cenário pode-se observar que o número de soluções aumenta de acordo com o tamanho da rede de estradas, fazendo com que se consumam quantidades extraordinárias de tempo para calcular uma solução. Sendo assim, o método de Simulated Annealing foi colocado em prática e apresentou soluções melhores que os métodos tradicionais, sendo eles: o TRANSYT (Traffic Network Study Tool), um simulador desenvolvido no laboratório de pesquisa de transportes de U.K.; e o método por algoritmos genéticos, que teve um melhor desempenho se comparado ao TRANSYT.

Outra possível técnica para a resolução de problemas combinatórios de otimização é o uso da rede neural Hopfield (Wang et al., 2002). O estudo avaliou a performance deste método simulando instâncias do problema do vendedor viajante de 10 cidades e também de outros métodos de validação de desempenho. A simulação demonstrou que este método foi capaz de gerar soluções 100% válidas consideradas ótimas ou quase ótimas.

Os modelos de otimização com abordagem de métodos heurísticos foram estudados na pesquisa de Guo et al. (2015) para serem aplicados no contexto da otimização do endereçamento de recursos para equipamentos de grande complexidade em empresas de manufatura, com foco nos recursos de alta qualidade. O problema em questão, denominado otimização combinatória de recursos de manufatura (OCRM), leva em conta 4 parâmetros para alimentar um algoritmo genético baseado em clusters e após implementado, foi comparado com outro algoritmo genético não baseado em clusters. Os resultados experimentais mostraram que o método proposto teve um melhor desempenho na busca pela solução ótima e também foi mais eficiente com a escalabilidade do problema.

Em anos recentes foi notado um crescimento de estudos com o interesse de integrar técnicas de aprendizado de máquina com meta-heurísticas para a resolução de problemas de otimização combinatória (Karimi-mamaghan et al. 2021). Entretanto como observado no estudo, a literatura não entrega uma análise técnica do assunto, criando a necessidade de realizar uma avaliação das tentativas de avanço no uso do aprendizado de máquina. O estudo em questão fornece uma visão técnica e compreensível da integração de técnicas de aprendizado de máquina em metaheurísticas para problemas envolvendo otimização combinatória.

Chassein et al. (2020) abordam problemas de otimização combinatória com cenários de custo K. Esse trabalho, ressalta o critério da média ponderada ordenada, comumente usada como agregadora de objetivos para computar uma solução, apontando a dificuldade de minimizar esse critério, fazendo com que o problema se torne NP-difícil, mesmo se o número de cenários for igual a 2. Tendo isso em vista, o trabalho propõe reduzir o número de cenários por meio de agregar os custos antes de resolver o problema.

Bengio et al. (2021) defendem em seu artigo o uso de aprendizado de máquina para a resolução de problemas combinatórios. O estudo ressalta que o aprendizado de máquina funciona como um candidato natural para a resolução desses problemas, sendo que com algoritmos e heurísticas feitas manualmente, demandariam uma grande quantidade de tempo e recursos para computar uma resolução. O artigo detalha uma metodologia para realizar a integração do aprendizado de máquina com os problemas propostos.

Mais relacionado ao contexto deste trabalho, Zajac e Huber (2021) levam em consideração os múltiplos objetivos que são estudados em uma operação logística, tais quais os problemas de rota, gerenciamento de frota de veículos, ou o clássico problema da menor rota. O trabalho tem como objetivo estruturar a literatura propondo uma classificação dos objetivos que foram agrupados em categorias e relacionados a diferentes elementos dos problemas.

Os métodos de clusterização são clássicos para a resolução de problemas de localização e definição de rotas, e foram estudados por Klose (1995) em sua pesquisa. Parte importante do problema se refere aos custos fixos e variáveis da operação, sendo que pelos custos variáveis, temos o transporte dos insumos das instalações para os pontos de demanda. Andreas ressalta que comumente se assume que os custos de transporte são proporcionais a distância entre os pontos de demanda e as instalações, mas isso deixa de ser verdade quando colocamos diversos consumidores na mesma rota de transporte. Sendo assim, ainda é crucial realizar o estudo para se chegar em um número próximo dos custos variáveis. A solução do problema se baseia em agrupar os pontos de demanda em clusters correspondentes as rotas e os anexar as instalações responsáveis por seu respectivo cluster.

Outro exemplo do uso de métodos de clusterização é mostrado no estudo de Kim et al. (2017), que utilizaram uma abordagem de cobertura para a solução do problema de localização de p-functional clusters. O processo de regionalização se baseia no método de delinear unidades geográficas em grupos tomando como base características similares ou atributos associados (Duque et al., 2007). Tendo isso em mente, é aceitável assumir que em um problema de localização, cada ponto irá participar exclusivamente de uma única região ou distrito. Entretanto, o trabalho de Kamyoun Kim nos mostra que essa afirmação é pouco realista quando tratamos de clusters funcionais como em áreas metropolitanas ou áreas de comércio, pois cada região não tem uma área exclusiva de cobertura com outras áreas.

3. Metodologia

Nesta sessão será apresentada inicialmente a classificação da pesquisa e as etapas metodológicas para o desenvolvimento do estudo. Na sequência é apresentada uma visão geral do algoritmo que foi desenvolvido para a obtenção da solução do problema, inclusive com os parâmetros empregados neste algoritmo. Na última subseção desta sessão são descritas as bases de dados que foram utilizadas nos experimentos que foram desenvolvidos.

3.1. Classificação da pesquisa e Etapas da Metodologia

A pesquisa desenvolvida neste projeto tem como característica ser uma pesquisa aplicada, já que trabalha com um problema eminentemente prático. A abordagem é quantitativa e sua finalidade é metodológica, pois seu objetivo é desenvolver um modelo computacional utilizando abstrações de técnicas de aprendizado de máquina e de modelagem matemática para resolver o problema de localização de instalações. Sobre os meios a utilizar no estudo, foi feita uma revisão bibliográfica do tema e da documentação das bibliotecas usadas.

O trabalho de pesquisa teve início por uma revisão bibliográfica e levantamento de bases de dados para os experimentos que seriam desenvolvidos. As bases utilizadas durante o desenvolvimento do projeto são públicas e referenciam a uma seleção das cidades mais proeminentes do Brasil, e dentre as informações de cada cidade que a base nos proporciona é utilizada a latitude, longitude e população de cada cidade. Outra base pública também utilizada no desenvolvimento do projeto, contém as cidades dos Estados Unidos e fornece a latitude, longitude e população de cada cidade. Foram definidas de forma mais detalhada e final as técnicas que seriam empregadas na solução do problema, e na sequência houve a necessidade de se definir ferramentas computacionais para implementação das técnicas e então, a fase de desenvolvimento e testes de algoritmos e de desenvolvimento de experimentos, onde foi possível a obtenção de resultados que puderam ser analisados. Finalmente, são feitos testes com diferentes bases de dados para comprovar a efetividade do modelo computacional desenvolvido.

Assim, pode-se dizer que a metodologia se desenvolveu segundo as etapas abaixo:

a)Revisão Bibliográfica

Neste passo, foram levantados trabalhos com o tema relacionado a esse projeto a fim de buscar referências para o problema de localização, técnicas de aprendizado de máquina, mais especificamente redes neurais artificiais e modelos de programação matemática.

b)Levantamento de Bases de Dados

Os experimentos utilizaram bases de dados (datasets) públicas disponíveis em sites científicos, procurando estudar em diferentes instâncias os resultados obtidos por uma solução obtida pela implementação de uma RNA do tipo SOM combinada com um modelo de programação binária.

c)Análise e detalhamento das técnicas a empregar no estudo

. Definiu-se o uso de técnicas de RNA do tipo SOM para resolver o problema de localização, obtendo o agrupamento dos pontos de demanda e consequentemente, as instalações. Em seguida, o modelo matemático de programação binária é utilizado a fim de otimizar a escolha feita pelo modelo de RNA.

d)Seleção das ferramentas computacionais adequadas ao desenvolvimento

Trata-se da seleção das ferramentas computacionais empregadas no desenvolvimento dos algoritmos necessários para a solução do problema e da apresentação de resultados. Para isto, foi definida a linguagem R

especializada em técnicas estatísticas e de ciência de dados. Em R utilizou-se as bibliotecas RSNNS para a normalização dos dados dos datasets, a biblioteca kohonen para gerar a grade e o modelo SOM, a biblioteca pracma para realizar o cálculo de distâncias utilizando coordenadas e as bibliotecas ompr e magrittr para a construção de um modelo matemático de programação binária.

e) Seleção das bibliotecas gráficas para visualização de resultados

Todas as análises de resultados foram tratadas com técnicas de visualização e com procedimentos estatísticos formais por meio de bibliotecas gráficas de plotagem de dados. Foi usada a biblioteca gráfica do R, a ggplot2 e a biblioteca gráfica R-Shiny. Esta foi usada para fazer a apresentação gráfica, no formato de um Diagrama Estrela (centroídes conectados a pontos de atendimento).

f) Desenvolvimento dos algoritmos

Nesta etapa teve-se dois tipos de desenvolvimento:

f1) Construção de uma RNA do tipo SOM para clusterizar os pontos de demanda e definir os protótipos (centroídes) dos clusters gerados, como sendo candidatos a instalações de atendimentos desses pontos de demanda;

f2) Construção de um modelo matemático de otimização para definir localizações ótimas a partir dos locais candidatos, e alocar pontos de demanda às instalações

g) Experimentos e Análise de Resultados

Inicialmente foram feitos testes com duas bases de dados diferentes, uma contendo as cidades dos Estados Unidos e a outra as cidades do Brasil, com sua latitude, longitude e população. Foi escolhida uma amostra dos datasets para que o estudo fosse direcionado a uma situação mais real a um problema de logística, sendo que para os Estados Unidos, são utilizados os dados das cidades do Estado de Delaware e para o Brasil, as cidades da região Sudeste e Nordeste. Nesta etapa, os algoritmos desenvolvidos puderam ser testados e gerar resultados, definindo localizações para instalações e a alocação de pontos de demanda a essas instalações. Após obter resultados com as duas primeiras bases de dados, foram selecionadas outras bases para realizar mais testes, a fim de comprovar que o algoritmo fosse efetivo na solução do problema de estudo do projeto.

3.2. Descrição do Algoritmo

Inicialmente são importadas as bibliotecas usadas para a criação do modelo SOM e para a normalização de dados.

Foram utilizadas duas bibliotecas do R: kohonen e RSNNS.

A função som() da biblioteca kohonen foi armazenada em uma variável de nome somFunc(), porque há outra função na biblioteca RSNNS com o mesmo nome, e assim, consegue-se diferenciar as duas funções. Se isto não for feito, o R studio sobrescreve a função de kohonen.

A seguir, a base de dados é importada e normalizada para o ambiente do algoritmo, contendo os pontos desejadas para o estudo e as variáveis de localização e demanda dos pontos.

Antes de dar início no modelo SOM, é necessário definir sua arquitetura que servirá como base para o modelo. Neste sentido, foi definida uma arquitetura com uma grade hexagonal, com dimensões de 3x4. E assim, o modelo é criado. Na implementação foi utilizada a função do R, renomeada `som()` da biblioteca `kohonen` e passando como parâmetros para a função: a base de dados já normalizada, a grade criada no passo anterior, o número de iterações realizadas, a taxa de aprendizado que foi deixada com o valor padrão, o argumento `keep.data`, indicando a intenção de armazenar as informações do modelo, e o raio que também é uma exigência do SOM.

O SOM gera protótipos, em que é possível pela propriedade `codes` do modelo, se obter os valores de cada um de seus atributos. Esses protótipos estão sendo considerados neste estudo, como sendo os centroides (centro geográfico) dos clusters gerados em cada neurônio do SOM (exemplares associados a cada neurônio). As coordenadas desses centroides correspondem, então, aos valores dos atributos desses protótipos.

Entretanto, como os pontos são passados ao modelo em sua forma normalizada, a localização dos centroides também é dada nesta forma e conseqüentemente, é necessário o uso de um método de normalização reversa. O método de normalização reversa é dado pelo `RSNNS` e recebe o vetor de valores que devem ser processados e os parâmetros de normalização já utilizados.

Uma vez geradas as instalações, é realizada a organização dos dados em datasets. Para os pontos de demanda, os dados são armazenados no data frame “`customer_locations`”, este contendo o índice `id`, latitude, longitude, população de cada cidade, e o centroide a qual a cidade está alocada, todos dados em sua forma não processada, ou seja, sem passar pelo método de normalização. Para os centroides, os dados são armazenados no data frame “`warehouse_locations`” contendo: o índice `id`; latitude e longitude, dados que já passaram pelo processo de normalização reversa; uma variável “`dist_to_mean`”, representando a distância entre a média da localização dos centroides para cada um deles; “`cost_per_square_meter`” representando o custo indicado para a localização do armazém, determinado por um vetor de distância entre os centroides e suas médias com o primeiro quartil representando o custo de 2000 reais por metro quadrado ($/m^2$), segundo quartil com 1500/ m^2 , terceiro quartil com 1000/ m^2 , e valores acima com 500/ m^2 ; “`total_population`” representando a população somada de todas as cidades cobertas por cada armazém; “`warehouse_size`”, dado em metros quadrados; custo de instalação do armazém, calculado multiplicando o custo por metro quadrado pelo tamanho do armazém.

Tendo os dados dos pontos de demanda e dos centroides definidos, a etapa de definir os custos fixos é completada sendo possível partir para a etapa da definição dos custos variáveis de transporte da operação. Para isso, é utilizada a biblioteca `pracma` contendo o método `haversine()`, este que recebe coordenadas geográficas de dois pontos e retorna a distância entre eles em quilômetros. O algoritmo prevê também o uso de distância euclidiana, substituindo o uso da função `haversine()`. É também utilizada a metodologia de cálculo de fretes proposta por Vallim Filho e Kurihara (2001) e pela

Agência Nacional de Transportes Terrestres (ANTT), na resolução número 5.820 que instituiu a Política de Preços Mínimos do Transporte Rodoviário de Cargas, e que foi emitida em decorrência da greve dos caminhoneiros ocorrida em maio de 2018. A seguir, é possível visualizar os procedimentos de cálculo de custos fixos (tabela 3.1) e custos variáveis (tabela 3.2) de um veículo. Estes custos definiram os parâmetros a serem utilizados no experimento. Os cálculos foram baseados na metodologia da ANTT (2018).

Os custos fixos do veículo são calculados em uma base mensal e são expressos em R\$/mês. Os custos variáveis do veículo são calculados em uma base quilométrica, sendo expressos em R\$/km. Todos os parâmetros necessários para o cômputo desses custos foram levantados no mercado, tais como: preços de veículos, preços de combustível, salários, etc. Para cômputo desses custos foi considerado um veículo médio (capacidade de 12t) cujas características e custos de insumos foram obtidos em fornecedores e portais do mercado de transportes de carga. A tabela 3.3 apresenta o resumo final desses custos.

Para o cálculo do custo de transporte é utilizado o custo final calculado na planilha da tabela 3.3 e é necessária também a função haversine da biblioteca *pracma* que recebe as coordenadas geográficas medidas em graus, e retorna a distância entre as duas em quilômetros. O valor é multiplicado pelo resultado do cálculo por km (tabela 3.3), que foi arredondado para R\$2,50, e pela volume do armazém em toneladas. A função que chamamos de *transportCostMatrixFact* realiza todos os cálculos e gera a matriz de custo de transporte. O custo de transporte total da operação considera toda a tonelagem que será movimentada em cada armazém e que deverá ser transportada.

Isto é feito pela seguinte equação:

$$CTr = Dc \cdot (CKM/Cap) \cdot (A \cdot h / \rho)$$

onde:

CTr = custo de transporte total em uma rota

Dc = distância do centróide (cidade) até o centróide médio

CKM = custo por km do veículo padrão de transporte

Cap = capacidade do veículo

A = área do armazém

h = pé direito do armazém

ρ = densidade média da carga

CUSTOS OPERACIONAIS DE VEÍCULOS			
PLANILHA ANTT			
Custos Fixos			
1. Reposição do Veículo			
Vr. do Veículo	230.000,00		
Valor Residual do Veículo	60%		
Perda durante a Vida Útil	92.000,00		
Vida Útil (meses)	72		
Custo Mensal de Reposição do Veículo	1.277,78		
2. Reposição do Equipamento de Carga			
Vr. do Equipamento de Carga	35.000,00		
Valor Residual do Equipamento de Carga	10%		
Perda durante a Vida Útil	31.500,00		
Vida Útil (meses)	84		
Custo Mensal de Reposição do Equipam.	375,00		
3. Remuneração Mensal de Capital			
Valor do Veículo	220.000,00		
Valor do Equipamento de Carga	35.000,00		
Vr. do 3o Eixo	10.000,00		
Total	265.000,00		
Taxa Anual de Juros	12%		
Custo Mensal de Capital	2.650,00		
4. Custos de Mão de Obra de Motorista			
Salário Nominal de Motorista	1.500,00		
Reserva de Encargos Sociais	90%		
No. de Motoristas por Veículo	1,5		
Custo Mensal de Motorista	4.275,00		
		5. Tributos sobre o Veículo	
		Vr. do Veículo	230.000,00
		% de IPVA	3%
		Custo Anual de IPVA	6.900,00
		DPVAT + Licenciamento	200,00
		Custo Anual (c/ DPVAT e Licenciamento)	7.100,00
		Custo Mensal de Tributos	591,67
		6. Riscos de Acidentes e Roubo do Veículo	
		Importância Segurada	230.000,00
		% Seguro do Veículo	3%
		Custo Anual	6.900,00
		Custo Mensal de Riscos do Veículo	575,00
		7. Riscos de Acidentes e Roubo do Equipamento de Carga	
		Importância Segurada	35.000,00
		% Seguro do Equipamento de Carga	6,5%
		Custo Anual	2.275,00
		Custo Mensal de Riscos do Equipamento	189,58
		Custo Fixo Mensal Total	9.934,03

Tabela 3.1. Tabela de Custos Fixos de Transporte.

CUSTOS OPERACIONAIS DE VEÍCULOS			
PLANILHA ANTT			
Custos Variáveis			
1. Manutenção			
Vr. do Veículo + Equipamentos	265.000,00		
Índice de Manutenção	1,0%		
Custo Mensal de Manutenção	2.650,00		
Km Média Mensal do Veículo	8.000		
Custo de Manutenção / Km	0,33		
2. Combustível			
Preço do Combustível (R\$/l)	3,00		
Rendimento (Km/l)	5,0		
Custo de Combustível / Km	0,60		
3. Lubrificantes			
Óleo de Motor			
Preço por Litro de Óleo	30,00		
Capacidade do Câter	14		
Litros por Reposição	1		
No. de Reposições	9		
Total de Litros (Troca + Reposições)	23		
Custo Total (Troca + Reposições)	690,00		
Km de Troca	10.000		
<i>Custo de Óleo de Motor / Km</i>	<i>0,069</i>		
Óleo de Transmissão			
Preço por Litro de Óleo	40,00		
Capacidade do Câmbio + Diferencial	20		
Custo Total	800,00		
Km de Troca	20.000		
<i>Custo de Óleo de Transmissão / Km</i>	<i>0,0400</i>		
Custo de Óleos / Km	0,109		
		4. Lavagem e Graxas	
		Preço da Lavagem	300,00
		Km entre Lavagens	4.000
		Custo de Lavagens / Km	0,08
		5. Pneu e Recauchutagens	
		Preço de um Pneu	1.000,00
		Preço de Câmara	75,00
		Preço de Protetor	25,00
		Preço Total de um Pneu	1.100,00
		Perda de Pneus (Acidentes)	3%
		Custo de Pneu com Perda	1.133,00
		Preço de Reforma de Pneu	250,00
		No. de Reformas por Pneu	2
		Custo de Reformas de Pneu	500,00
		No. de Câmaras Adicionais por Pneu	1
		Custo de Câmaras Adicionais por Pneu	75,00
		No. de Protetores Adicionais por Pneu	1
		Custo de Protetores Adicionais por Pneu	25,00
		Custo Total por Pneu	1.733,00
		No de Pneus por Veículo	11
		Custo Total de Pneus por Veículo	19.063,00
		Vida Útil de Pneus	150.000
		Custo de Pneus / Km	0,127
		Custo Variável Total / Km	1,24

Tabela 3.2. Tabela de Custos Variáveis de Transporte.

Km por Mês	8.000
C. Variável por Km	1,24
C. Variável Mensal	9.938,69
C Fixo Mensal	9.934,03
C. Mensal TOTAL : Fixo + Variável	19.872,72
C. KM TOTAL : Fixo + Variável	2,48

Tabela 3.3. Resumo dos Custos Fixos e Variáveis do Veículo.

O passo seguinte se dedica a criar uma matriz utilizada no modelo de programação matemática, contendo o custo de transporte de cada centroide para cada cidade. O modelo de programação é criado nesse momento, utilizando como variáveis: x_{ij} do tipo binário, representando se uma cidade “i” está atrelada a um centroide “j”; e a variável y_j também do tipo binário, representando se o centroide “j” será selecionado para receber uma instalação.

$$\text{Min } [\sum_i \sum_j \text{CustoTransporte}_{i,j} \cdot x_{i,j} + \sum_j \text{CustoFixo}_j \cdot y_j] \quad (3.1)$$

Sujeito a:

$$\sum_j x_{ij} = 1 ; i=1, \dots, n ; j=1, \dots, m \quad (3.2)$$

$$x_{ij} \leq y_j ; i=1, \dots, n ; j=1, \dots, m \quad (3.3)$$

$$x_{ij} \in \{0,1\} ; i=1, \dots, n ; j=1, \dots, m; y_j \in \{0,1\} ; j=1, \dots, m \quad (3.4)$$

O modelo é resolvido por meio do “solver” glpk da infraestrutura de otimização do R (R Optimization Infrastructure – ROI).

3.3. Aplicação do algoritmo em bases de dados públicas

As bases de dados utilizadas são bases públicas propostas por Uchoa et al. (2017) contendo 100 conjuntos variando entre 101 e 1001 pontos, sendo que dessas bases, utilizamos apenas 10. Elas são representativas, e atendem ao objetivo de... porém não correspondem a nenhuma operação logística real em particular. Cada ponto nas bases é representado pelas variáveis c_x e c_y , representando a localização de cada ponto em um plano, e a variável $quantity$, representando o nível de demanda do ponto.

Por c_x , c_y e $quantity$ serem dados em valores de escalas diferentes à base de dados usada no desenvolvimento do algoritmo, foi necessário realizar alterações nos cálculos para que o modelo matemático encontrasse a melhor solução. Para o cálculo dos custos de transporte, houve a substituição da função `haversine()` pelo cálculo euclidiano de distâncias pois a função usada originalmente recebia como parâmetro as coordenadas geográficas dos dois pontos retornando o valor em quilômetros, e os valores de c_x e c_y variam entre os valores de 1 a 1000, não se encaixando no antigo padrão. Também foi necessário alterar o cálculo para a definição do tamanho das instalações que originalmente eram definidas pelo tamanho da população dividida por 100, ou seja, 0,01 metros por habitante. Agora, consideramos 1 metro por habitante

visto que o valor de quantity também não é expressivo como nas bases de dados originais.

4. Resultado e Discussão

Esta sessão apresenta os resultados do estudo, inicialmente introduzindo as saídas gráficas do algoritmo aplicado à base de dados brasileira, utilizada no desenvolvimento do algoritmo, seguindo para as principais saídas gráficas das bases de dados de Uchoa et al. (2017).

Tabela 4.1 - Resultados dos Experimentos

BASE DE DADOS		SOLUÇÃO ÓTIMA	
Nome	No. de Pontos	No de Instalações Geradas	Custo da Instalação
Base brasileira	211	8	115757555
X-n101-k25	101	6	631500
X-n204-k19	204	9	9830000
X-n303-k21	303	5	2238500
X-n411-k19	411	8	1785500
X-n524-k137	524	5	2517000
X-n613-k62	613	5	4123500
X-n716-k35	716	5	6840500
X-n819-k171	819	7	33785500
X-n916-k207	916	9	5998500
X-n1001-k43	1001	9	3711000

O algoritmo desenvolvido foi capaz de se chegar em um custo mínimo para a operação das diferentes bases de dados utilizadas. A título de ilustração apresenta-se a figura 4.1 com um gráfico mostrando a seleção de 211 cidades da base de cidades brasileiras, sendo representadas pelos pontos pretos, e 8 instalações sendo representadas pelos triângulos vermelhos.

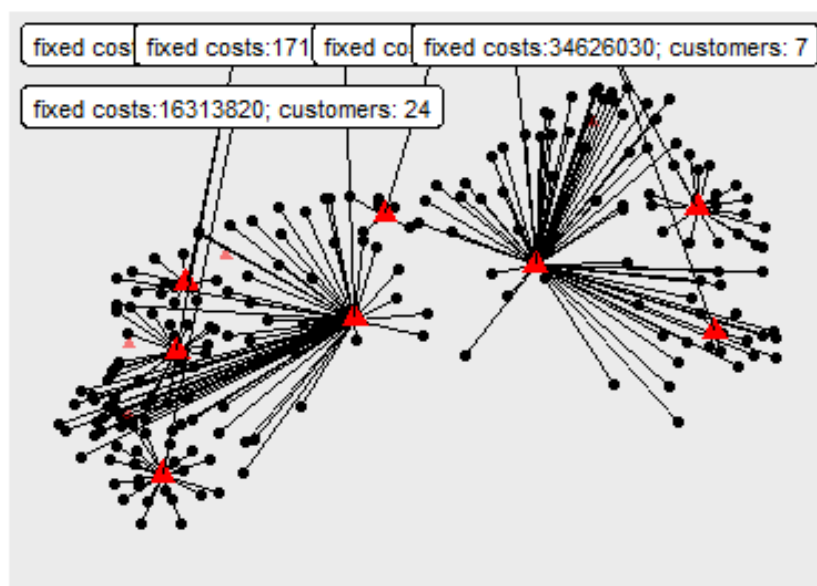


Figura 4.2. Alocação de instalações para as cidades brasileiras.

Para a faixa de 100 pontos (figura 4.2), o algoritmo determinou 6 instalações, sendo notável a presença de pontos que foram alocados a instalações distantes de sua localização.

Entretanto, a medida que o número de pontos aumenta, o algoritmo realiza uma seleção mais precisa, sendo até possível distinguir regiões ou clusters, como podemos ver nos gráficos para as bases de 500 pontos (imagem 4.4) e 1000 pontos (imagem 4.5). É possível observar no gráfico de 300 pontos (imagem 4.3), tendo 5 instalações definidas, que o algoritmo se adapta a regiões distantes do centro de concentração de pontos, alocando instalações para cobrir a necessidade destes mais distantes.

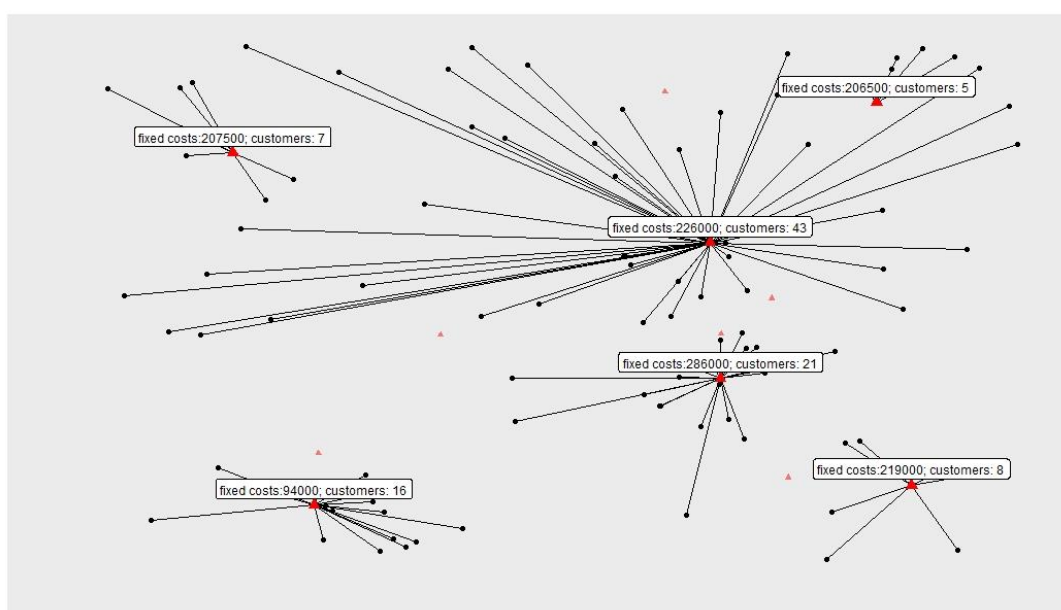


Figura 4.3. Resultado gráfico para 100 pontos – Base X-n101-k25.

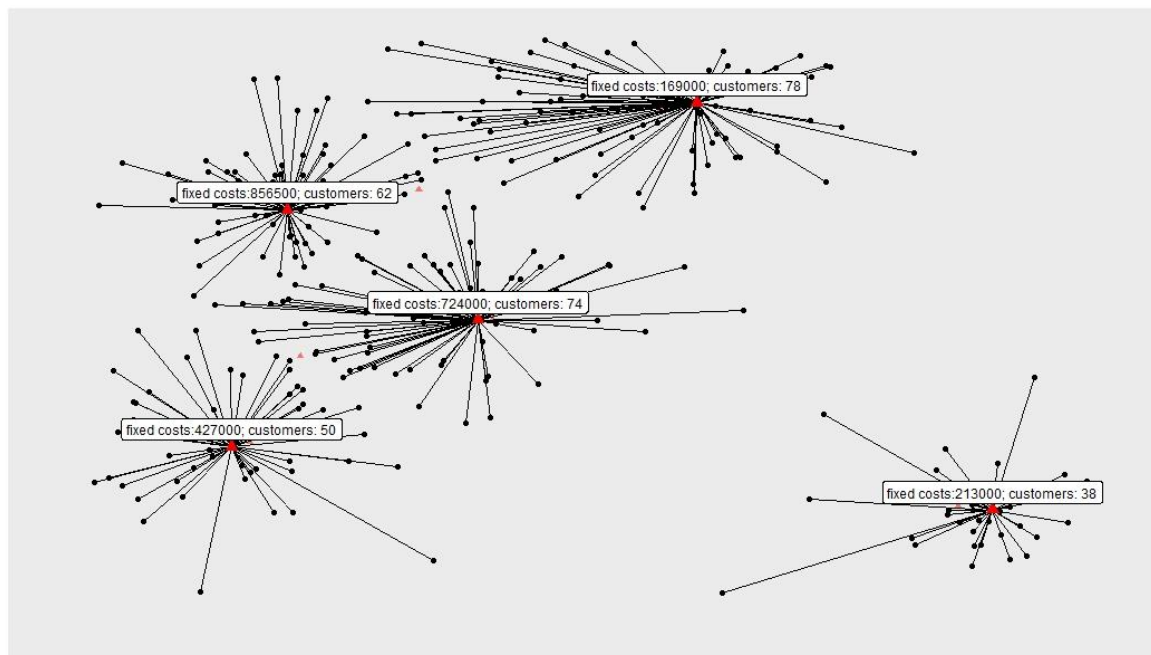


Figura 4.4. Resultado gráfico para 300 pontos – Base X-n303-k21

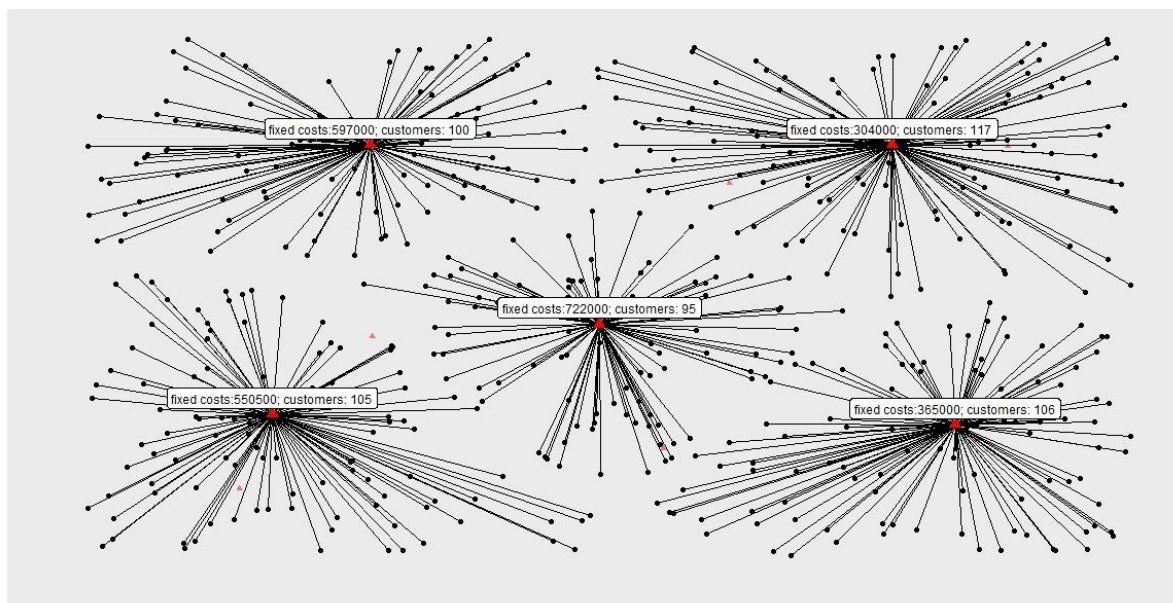


Figura 4.5. Resultado gráfico para a base de 500 pontos – Base X-n524-k137.

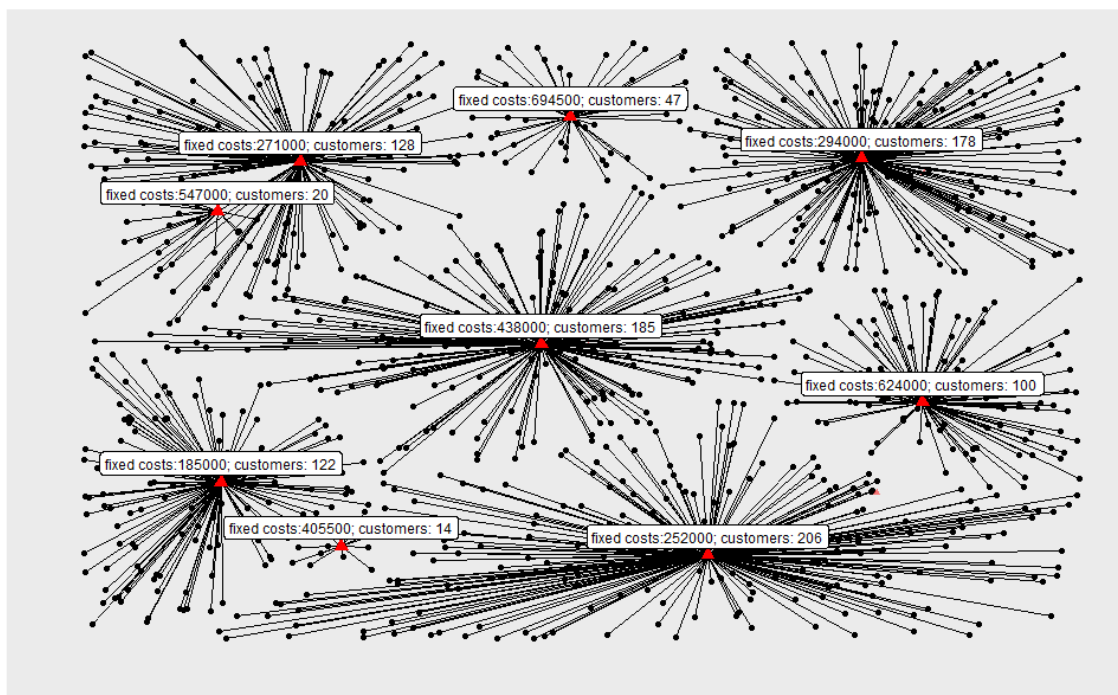


Figura 4.6. Resultado gráfico para a base de 1000 pontos – Base X-n1001-k43.

5. Considerações Finais

O projeto teve sucesso em buscar o retorno gráfico e o cálculo dos dados referentes aos armazéns gerados artificialmente pela rede neural SOM tal qual testar sua eficiência com novas bases de dados de diferentes faixas de pontos com valores em escalas diferentes. Foi possível se chegar em um custo mínimo da operação definindo a quantidade de instalações, otimizando a escolha por meio do método de modelagem matemática. Como apontado pelo trabalho de Brandeau e Chiu (1989), a pesquisa mostra uma grande relação de aplicações, nas áreas privada e pública, e pode ser reaproveitada para um conjunto de novos trabalhos e pesquisas, visto que o projeto opta por uma forma entre várias de se buscar a melhor localização de instalações.

É possível desenvolver novos estudos para otimizar o algoritmo, buscando comparar o método utilizado com outras técnicas alternativas de metaheurísticas ao modelo matemático. Dessa forma seria possível novamente comprovar a eficiência do algoritmo ou sua falta.

Sugere-se realizar estudos relacionados ao problema de localização, tanto aplicando o algoritmo desenvolvido neste trabalho a novas bases de dados, como com diferentes contextos, como por exemplo: redes de saúde, como hospitais, que precisam atender a demanda de acidentes enviando ambulâncias; rede de atendimento policial, no qual há uma instalação que atende uma área enviando viaturas.

Referências

- SAMUEL WILLIAN ALVES WU, ARNALDO RABELLO DE AGUIAR BALLIM FILHO (2020). O Problema Combinatório de Localização de Instalações. XVI Jornada de Iniciação Científica e X Mostra de Iniciação Tecnológica – 2020, Universidade Presbiteriana Mackenzie, Faculdade de Computação e Informática, São Paulo, SP.
- ANTT (2019). RESOLUÇÃO nº 5.849, de 16 de julho de 2019. Estabelece as regras gerais, a metodologia e os coeficientes dos pisos mínimos, referentes ao quilômetro rodado na realização do serviço de transporte rodoviário remunerado de cargas, por eixo carregado, instituído pela Política Nacional de Pisos Mínimos do Transporte Rodoviário de Cargas - PNPM-TRC.
- BALLOU, R.H. (2006) Gerenciamento da cadeia de suprimentos/ Logística empresarial, 5 ed., Bookman Editora. Porto Alegre, RS.
- BECCENERI, J.C., (2012) Metaheurísticas e Otimização Combinatória: Aplicações em Problemas Ambientais. 2 ed..
- BRANDEAU, M. L., CHIU, S.S. (1989). “An overview of representative problems in location research”; Management Science, vol. 35 , n.6, p.645-674.
- COLORNI, A., DORIGO, M., MAFFIOLI, F., MANIEZZO, V., RIGHINI, G., TRUBIAN, M. (1997). “Heuristics from nature for hard combinatorial optimization problems”; International Transactions in Operational Research, No.3.1 ,p. 1 - 38.
- CRAINIC, T.G. e LAPORTE, G. (1997) “Planning models for freight transportation. ” European Journal of Operational Research 97, p. 409-438.
- DULLAERT, W.; VERNIMMEN, B.; WITLOX, F. (2002) A hybrid approach to designing inbound-resupply strategies. v.20.
- GITMAN, L. J. (2010). Princípios de Administração Financeira. 12ª Edição. Pearson Universidades. São Paulo, SP. 800p.
- HAYKIN, S. Neural Networks and Learning Machines. 3ª ed, Editora Prentice-Hall, 2008.
- KLOSE, A.; DREXL, A. “Facility location models for distribution system design. (2015) ” European Journal of Operational Research. 162 - 4–29.
- LUKE, S. Essentials of Metaheuristics. 2 ed, p. 9-12, 2015
- MARTINS, E. (2018). Contabilidade de Custos - 11ª Edição. Editora Atlas. São Paulo, SP. 408p.
- MATSUDA, A.; YOSHIMOTO, K. (2003). On the use of neural networks to solve location problems: Application to the long-term and chronological model. DSI & APDSI, Shanghai
- MICHALEWICZ, M.; SCHMIDT, M.; MICHALEWICZ, Z.; CHIRIAC, C. (2005) Case study: an intelligent decision support system. v.20.
- NETO, L. B.; BECCENERI, J. C.; SILVA, J. D. S.; LUZ, E. F. P.; NETO, A. J. S. (2010) Fundamentos de Otimização e Inteligência Artificial. n. 4, p. 35-37.

NTC & Logística. (2014) Manual de cálculo de custos e formação de preços do transporte rodoviário de cargas – 2014. Associação Nacional do Transporte de Cargas e Logística. São Paulo, SP.

OWEN, S.H. e DASKIN, M. S. (1998) “Strategic facility location: a review. ” European Journal of Operational research. No. 111, 423-447.

PIRLOT, M. (1996). “General local search methods “. European Journal of Operational Research. No. 92, p.493-511.

ReVELLE, C.S.; EISELT, H.A.; DASKIN, M.S. (2008) A bibliography for some fundamental problem categories in discrete location science. European Journal of Operational Research 184: 817–848.

SILVA, L. A.; PERES, S. M.; BOSCARIOLI, C. (2016). “Introdução à Mineração de Dados: Com Aplicações em R”. 1ª Ed. Elsevier Editora Ltda. R.J. 277 p.

SILVA, M. R. ; CUNHA, C. B. (2017). A tabu search heuristic for the uncapacitated single allocation p-hub maximal covering problem. European Journal of Operational Research 262, (3), 954-965

VALLIM FILHO, A. R. A., KURIHARA, T. (2001). Um modelo matemático de decisões estratégicas para a logística de distribuição de produtos. Revista Mackenzie de Engenharia e Computação: v.2, p.42 - 62. São Paulo, SP.

VALLIM Fo , A. R. A. (2004) Localização de centros de distribuição de carga: contribuições à modelagem matemática. Tese de Doutorado, Escola Politécnica da Universidade de São Paulo. São Paulo, 286p..

VENABLES, W. N.; SMITH, D. M. and the R Core Team. (2016). An Introduction to R. Notes on R: A Programming Environment for Data Analysis and Graphics. Version 3.3.2 (2016-10-31)

Bell, J. E. (2003) “A simulated annealing approach for the composite facility location and resource allocation problem: a study of strategic positioning of US air force munitions”. Doctoral Thesis. Auburn University. 166p.

CHIYOSHI, F. Y. ; GALVAO . A Statistical Analysis of Simulated Annealing Applied to the p-Median Problem. Annals of Operations Research, v. 96, p. 61-74, 2000.

SUE ABDINNOUR-HELM. (2001). Using simulated annealing to solve the p-Hub Median Problem. International journal of physical distribution & logistics management : IJPD & LM. - Bradford : Emerald, ISSN 0020-7527, ZDB-ID 1842791. - Vol. 31.2001, 3, p. 203

STEVEN J. D’AMICO, SHOUU-JIUN WANG, RAJAN BATTA, CHRISTOPHER M. RUMP. (2002). A simulated annealing approach to police district design. Blasland, Bouck & Lee, Inc. 6723 Towpath Road, Syracuse, NY 13214, USA; First USA Bank, DE1-1027, 201 North Walnut Street, Wilmington, DE 19801, USA; Department of Industrial Engineering, University at Buffalo, State University of New York, 342 Lawrence D. Bell Hall, Box 602050 Buffalo, NY 14260, USA

S. L. HAKIMI (1964). "Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph". *Operations Research*, Vol. 12, No. 3 (May - Jun., 1964), pp. 450-459

LIXUN ZHANG, GERARD RUSHTON (2006). Optimizing the size and locations of facilities in competitive multi-site service systems. Department of Geography, The University of Iowa, Iowa City, IA 52242, USA

SHAMS-UR RAHMAN, DAVID K. SMITH (1999). Use of location-allocation models in health service development planning in developing nations. The Graduate School of Management, The University of Western Australia, Nedlands, WA 6907, Australia School of Mathematical Sciences, University of Exeter, North Park Road, Exeter, EX4 4QE, UK

TOSHIHIKO ODA, TORU OTOKITA, TOMOMITSY TSUGI, YOSHITO MASHIYAMA (1997). Application of Simulated Annealing to optimization of Traffic Signal Timings. Information Systems Division, Matsushita Communication Industrial Co., Ltd. Matsushita Communication SeruJai Laboratory Co., Ltd.

RONG LONG WANG, ZHENG TANG, QI PING CAO (). A learning method in Hopfield neural network for combinatorial optimization problem. Faculty of Engineering, Toyama University, Toyama-shi, 930-8555, Japan Tateyama Systems Institute, Toyama-shi, 930-0001, Japan

SHUNSHENG GUO, BAIGANG DU, ZHAO PENG, XIAORONG HUANG, YIBING LI (2015). Manufacturing resource combinatorial optimization for large complex equipment in group manufacturing: A cluster-based genetic algorithm. School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China. School of Management & Economics, Hubei Polytechnic University, Huangshi 435003, China

MARYAM KARIMI-MAMAGHAN, MEHRDAD MOHAMMADI, PATRICK MEYER, AMIR MOHAMMAD KARIMI-MAMAGHAN, EL-GHAZALI TALBI (2020). Machine learning at the service of meta-heuristics for solving combinatorial optimization problems: A state-of-the-art. IMT Atlantique, Lab-STICC, UMR CNRS 6285, Brest F-29238, France Department of Electrical and Computer Engineering, University of Tehran, Tehran, Iran Department of Computer Science, University of Lille, CNRS UMR 9189, Centre de Recherche en Informatique Signal et Automatique de Lille (CRISTAL), Lille F-59000, France

ANDRÉ CHASSEIN, MARC GOERIGK, ADAM KASPERSKI, PAWEŁ ZIELINSKI (2020). Approximating combinatorial optimization problems with the ordered weighted averaging criterion. Data Analytics Center of Excellence, Deutsche Post DHL Group, Bonn, Germany Network and Data Science Management, University of Siegen, Germany Faculty of Computer Science and Management, Wrocław University of Science and Technology, Poland Faculty of Fundamental Problems of Technology, Wrocław University of Science and Technology, Poland

YOSHUA BENGIO, ANDREA LODI, ANTOINE PROUVOST (2020). Machine learning for combinatorial optimization: A methodological tour d'horizon. Canada Excellence Research Chair in Data Science for Decision Making, École Polytechnique

de Montréal, Pavillon André-Aisenstadt 2920, Chemin de la Tour Montreal, Qc, H3T 1J4 Canada; Mila, Institut Québécois d'Intelligence Artificielle, Pavillon André-Aisenstadt 2920, Chemin de la Tour Montreal, Qc, H3T 1J4 Canada; Université de Montréal, Département d'Informatique et de Recherche Opérationnelle, Pavillon André-Aisenstadt 2920, Chemin de la Tour Montreal, Qc, H3T 1J4 Canada

SANDRA ZAJAC, SANDRA HUBER (2020). Objectives and methods in multi-objective routing problems: a survey and classification scheme. Ruhr University Bochum, Universitätsstraße 150, Bochum 44801, Germany; Helmut Schmidt University, Holstenhofweg 85, Hamburg 22043, Germany

ANDREAS KLOSE (1995). Using Clustering Methods in Problems of Combined Location and Routing. Institut für Unternehmensforschung (Operations Research), Hochschule St. Gallen Bodanstr. 6, 9000 St. Gallen, Switzerland

KAMYOUNG KIM, YONGWAN CHUN, HYUN KIM (2017). p-Functional Clusters Location Problem for Detecting Spatial Clusters with Covering Approach. Department of Geography Education, Kyungpook National University, Daegu, South Korea, 2 School of Economic, Political and Policy Sciences, University of Texas at Dallas, Richardson, TX, USA, 3 Department of Geography, University of Tennessee, Knoxville, TN, USA

DUQUE, J. C., RAMOS, R., and SURINACH, J. (2007). "Supervised Regionalization Methods: A Survey." *International Regional Science Review* 30(3), 195–220.