O Problema combinatório de localização de instlações

Samuel Willian Alves Wu e Arnaldo Rabello de Aguiar Vallim Filho

**Apoio:**Escolher um item.

RESUMO

Este projeto busca o desenvolvimento de uma suíte de heurísticas e metaheurísticas para solução de problemas de otimização combinatória de diferentes tipos. O projeto já teve uma Fase I em que foram desenvolvidos algoritmos para problemas de definição de rotas, para atendimento de pontos de demanda em uma dada região geográfica. Nesta fase II, o projeto se concentra em problemas de “clusterização” e problemas de determinação da localização de instalações (centroides) para atendimento de pontos de demanda.

No caso específico deste projeto, o trabalho pretende contemplar o Problema de “Localização de Instalações”, que é um problema combinatório clássico da área de Mineração de Dados (Silva et al., 2016), em que as instalações são representadas por centroides. Do ponto de vista prático, uma de suas aplicações está associada à Logística, e representa a mais importante decisão estratégica de uma operação Logística (Ballou, 2006). Neste projeto, pretende-se resolver este problema aplicado à área de Logística, por meio de técnicas de redes neurais artificiais.

Palavras-chave: Digite seu texto contendo 3 palavras

ABSTRACT

This project seeks the development of heuristic and metaheuristic suits for solutions of combinatory otimization problems of diferente types. The project had a phase 1, witch was developed algorithms for route definition problems, for servisse of demand points in a especific geografic region. In the phase 2, the project focus in clusterization problems and the definition of centroids locations to attend demand points.

In the especific case of this project, the work contemplate the problem of “centroid locations”, witch is a classic data mining combinatory problem (Silva et al., 2016), whereby the instalations are represented with centroids. From the pratical point of view, one of the applications is associated to logistic, and represents the most important strategic decision of a logistic operation (Ballou 2006). In this project, is intended to resolve this problem applied to the logistic area, with neural networks techniques.

Keywords: Digite seu texto contendo 3 palavras

1. INTRODUÇÃO

Este trabalho é parte de um projeto mais amplo denominado “Suíte de Soluções de Problemas Combinatórios – Fase II, que está sendo conduzido no laboratório de pesquisa BigMAAp – Big Data e Métodos Analíticos Aplicados, da Faculdade de Computação e Informática – FCI. Este projeto busca o desenvolvimento de uma suíte de heurísticas e metaheurísticas para solução de problemas de otimização combinatória de diferentes tipos. O projeto já teve uma Fase I em que foram desenvolvidos algoritmos para problemas de definição de rotas, para atendimento de pontos de demanda em uma dada região geográfica. Nesta fase II, o projeto se concentra em problemas de “clusterização” e problemas de determinação da localização de instalações (centroides) para atendimento de pontos de demanda.

No caso específico deste projeto, o trabalho contempla o Problema de “Localização de Instalações”, que é um problema combinatório clássico da área de Mineração de Dados (Silva et al., 2016), em que as instalações são representadas por centroides. Do ponto de vista prático, uma de suas aplicações está associada à Logística, e representa a mais importante decisão estratégica de uma operação Logística (Ballou, 2006). Neste projeto, pretende-se resolver este problema aplicado à área de Logística, por meio de técnicas de redes neurais artificiais.

***1. 1 Problema de Pesquisa***

Considerando-se o problema da localização no contexto de uma operação logística, pode-se dizer que de forma básica, o problema contempla um conjunto de pontos distribuídos em uma área geográfica, que serão atendidos em suas necessidades por instalações logísticas. Cada ponto tem uma demanda a ser satisfeita (medida em alguma unidade de peso, volume ou outra unidade) e cada instalação deverá atender um conjunto desses pontos.

Essas instalações logísticas podem ser:

. Unidades de Produção (fábricas);

. Centros de Armazenagem de Matéria Prima;

. Centros de Armazenagem de Produto Acabado;

. Centros de Distribuição;

. Operadores Logísticos;

. Outros

A literatura, por outro lado, apresenta diversos tipos de problemas de localização (*facility location problem*) e alguns autores apresentaram taxonomias para identificar os tipos e características desses problemas, como Klose e Drexl (2005) e ReVelle, Eiselt e Daskin (2008).

De forma geral, pode-se dizer que uma vez que a função da instalação esteja definida, a questão básica de localização de instalações envolve determinar:

Quantidade de instalações;

Localizações das instalações;

Alocação de pontos de demanda a cada instalação;

Alocação de fontes que irão abastecer cada instalação.

Nem sempre o foco do problema envolve todas essas decisões. Há problemas, por exemplo, que não consideram o abastecimento das instalações, como é o caso de postos de saúde ou postos de corpo de bombeiros. Já se as instalações forem, por exemplo, diferentes filiais de uma indústria, o abastecimento dessas Unidades de Produção, com matéria prima, a partir de fornecedores (fontes de fornecimento) é ponto chave a ser definido.

No caso específico deste projeto pretende-se trabalhar com um problema de localização que envolva as quatro decisões apresentadas acima.

Na maior parte dos casos, o problema da localização é resolvido tendo-se por objetivo a minimização de um índice que multiplica a distância percorrida pela quantidade de carga movimentada, ou então, pela minimização do custo logístico da operação.

Note-se que este é um problema de otimização combinatória, pois envolve uma grande quantidade de alternativas de solução (soluções viáveis), por isso, as técnicas de solução, em geral envolvem, metaheurísticas, já que este é um problema NP-Difícil, e assim, problemas de grande porte não podem ser resolvidos por algoritmos exatos, pois o tempo computacional cresce exponencialmente.

Nesta pesquisa pretende-se trabalhar com uma abordagem de solução baseada em redes neurais o que não é uma abordagem tão comum, já que a maioria das soluções propostas para esse tipo de problema se baseia em metaheurísticas. E na sequência, uma vez obtida a solução “ótima” do problema, propor uma apresentação gráfica para essa solução.

Assim, a questão que se coloca na pesquisa é:

“*como se comporta a solução de um problema da localização quando resolvido por uma rede neural artificial (RNA), quais as dificuldades e facilidades de implementação dessa técnica para solucionar esse tipo de problema, e como pode ser implementada uma apresentação gráfica da solução obtida por RNA?”*

***1. 2 Justificativa***

A importância do estudo de localizações é mostrada pelo grande número de aplicações que são encontradas em muitas áreas de atividade. O trabalho de Brandeau e Chiu (1989) ilustra essa aplicabilidade, com uma pesquisa em que mostra uma grande relação de aplicações, nas áreas privada e pública. Exemplos da área privada são: 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. 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.

Crainic e Laporte (1997) mostram a importância estratégica da localização esclarecendo que as decisões estratégicas determinam, na verdade, o desenvolvimento geral das políticas de uma empresa e formatam de modo abrangente as estratégias operacionais do sistema sobre um relativamente longo horizonte de tempo e, dentre estas decisões estratégicas, está a localização de instalações.

Owen e Daskin (1998) afirmam que a localização de instalações é um aspecto estratégico crítico em uma grande variedade de atividades públicas e privadas e acrescentam ainda, que a decisão sobre uma nova instalação é tipicamente um projeto sensível ao tempo e ao custo. Localizações de instalações são definidas para permanecer em operação por um longo tempo. É assim, portanto, decisão estratégica vital.

Ballou (2006) afirma que a localização de instalações pode ser considerada como o mais importante problema de planejamento logístico estratégico em um grande número de empresas. É esta decisão que estabelece as condições para a seleção apropriada e a boa administração dos serviços de transporte e dos níveis de estoque.

A importância da questão é, assim, bem reconhecida na literatura.

Sobre as abordagens de solução, os trabalhos mostram que na grande maioria dos casos a solução não se dá por técnicas de Inteligência Artificial, apesar de ser esta uma das áreas que mais cresce em computação (Vallim Filho, 2004).

Tratar assim o problema com uma abordagem diferenciada, em particular, com técnicas de uma área (Inteligência Artificial) com grande destaque e crescimento no momento, pode trazer contribuições importantes para essa área do conhecimento.

***1. 3 Objetivo***

O objetivo geral deste projeto é o desenvolvimento de um conjunto de experimentos para estudar o comportamento das soluções geradas para diferentes instâncias de um problema de localização.

A ideia básica da pesquisa é testar um tipo de solução com pouca utilização na literatura, que é a abordagem por RNA.

Em termos de objetivos específicos, pretende-se:

. Desenvolver experimentos com diferentes instâncias utilizando-se bases de dados clássicas;

. ajustar e treinar um algoritmo de otimização baseado em Redes Neurais Artificiais (RNA) para solução do problema da localização;

. comparar as soluções obtidas por meio de visualizações e análises estatísticas com apoio de uma biblioteca gráfica.

Em termos de ferramentas, pretende-se trabalhar com a linguagem R (Venables et al., 2016) e sua biblioteca gráfica R Shiny. Durante a pesquisa será avaliada a viabilidade e necessidade de uso do Python e de seu ambiente computacional, com suas bibliotecas matemáticas, como mais uma alternativa, para implementação do algoritmo de RNA.

1. REFERENCIAL TEÓRICO

Este capítulo tem como objetivo apresentar trabalhos científicos que possuem relação com o tema do presente projeto. Os materiais estudados auxiliaram no entendimento do problema e no conhecimento das estratégias de solução com o objetivo de tentar tratar a questão de localização de instalações.

O problema da localização tem sido resolvido por modelo exatos de Programação Matemática (PM) e por metaheurísticas (Vallim Filho, 2004). A PM é empregada para problemas de porte pequeno e médio, mas problemas de grande porte, que são os mais frequentes na prática são resolvidos por metaheurísticas, já que conforme dito acima, o problema é NP-Difícil.

O termo “metaheurística” deriva da composição de duas palavras gregas: “heurística” termo associado a um conhecimento circunstancial e o prefixo “meta” que significa “após”, indicando uma camada superior do processo de descoberta de uma solução para um problema. (Neto et al. 2010). Nesse sentido, metaheurísticas são técnicas que abrangem classes mais amplas de problemas. Assim, por exemplo: uma mesma metaheurística que pode ser usada para resolver problemas de localização de instalações, pode também resolver problemas de roteirização de veículos. Uma metaheurística, na verdade, não é um método pronto para solucionar um problema. Estes métodos 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 (Pirlot, 1996). Uma metaheurística pode ser considerada assim, 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).

Tratam-se de técnicas que estão na fronteira entre pesquisa operacional e inteligência artificial (Colorni et al., 1997).

Há uma quantidade grande de metaheurísticas. Laporte (2009) apresenta as principais categorias:

. metaheurísticas de busca local, tais como: *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: redes neurais e colônia de formigas.

Especificamente sobre o problema da localização, Vallim Filho (2004) mostrou que a localização ótima de Centros de Distribuição (CD) pode ser determinada através da metaheurística *Simulated Annealing* (SA). Para isto foi definida como função objetivo que minimiza os custos de localização e de transporte em uma rede com dois estágios transferência para CD’s e entrega nos destinos finais.

Pirlot (1996) apresenta um tutorial para três metaheurísticas amplamente utilizadas: SA, busca tabu e algoritmos genéticos. Uma descrição e aplicações das técnicas são apresentadas no artigo. Colorni et al. (1997) descreveram os principais aspectos das metaheurísticas derivadas da natureza, incluindo uma revisão e uma aplicação de SA. D'Amico et al. (2002) considerou o problema de redistritamento ou redesenho dos limites do comando policial através de metaheurística, e por conseguinte ficariam definidas as localizações das delegacias.

Bell (2003) desenvolveu uma pesquisa para construir um modelo para fornecer respostas sobre a melhor forma de reposição de inventários de munições da Força Aérea dos EUA necessários para futuros conflitos para uma variedade de cenários de demanda. Chiyoshi e Galvão (2000) desenvolveram uma análise estatística do SA aplicado ao conhecido problema da p-mediana, em que a mediana é a localização da instalação, a fim de identificar o desempenho da técnica. Eles testaram um conjunto de problemas, executando cada problema com no máximo cem fluxos diferentes de números aleatórios, o que é necessário para aplicar a metaheurística. Outro artigo interessante relacionado ao tema é o de Abdinnour-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 SA.

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 (neurôinios), como apontado por Leandro Augusto (2016). De forma auto-organizada, 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.

1. METODOLOGIA

A pesquisa desenvolvida neste projeto é um tipo de pesquisa aplicada, já que trabalha com um problema eminentemente prático.

A abordagem é quantitativa, e sua finalidade é metodológica, pois o objetivo foi desenvolver um modelo computacional. Sobre os meios a utilizar no estudo, foi feita uma revisão bibliográfica do tema e da documentação das bibliotecas usadas. Os experimentos utilizaram bases de dados (*datasets*) disponíveis em sites científicos, procurando estudar em diferentes instâncias os resultados obtidos por uma solução obtida pela implementação de uma RNA do tipo SOM.

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, utilizamos os dados das cidades do Estado de Delaware, e para o Brasil, escolhemos as cidades da região Sudeste e Nordeste.

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.

As ferramentas computacionais a empregar no estudo foram a linguagem R especializada para trabalhos com estatística e de ciência de dados e a biblioteca gráfica do R, a ggplot2. O problema da localização foi resolvido por Redes Neurais, em R utilizando 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 do modelo matemático de programação linear. Procedemos à leitura da solução pelo R-Shiny para fazer a apresentação gráfica, no formato de um Diagrama Estrela (centroides conectados a pontos de atendimento).

Importamos as bibliotecas usadas para a criação da grade e do modelo SOM, e para a normalização de dados. Guardamos a função som() da biblioteca kohonen em uma variável de nome somFunc(), pois a biblioteca RSNNS também possui uma função com mesmo nome, e ao importa-la, o R studio sobrescreve a função de kohonen.

A seguir, importamos a base de dados que está no formato csv para o ambiente do algoritmo, reunimos as cidades desejadas para o estudo e extraímos delas as variáveis de latitude, longitude e população para normalizarmos os dados, os guardando em um data frame *data\_train\_matrix*. Então criamos a grade do tipo hexagonal de dimensões 3x4. O modelo som é criado a seguir, utilizando a função da biblioteca kohonen *som()*, que renomeamos para *somFunc().* Passamos como parâmetro 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 deixamos com o valor padrão, o argumento keep.data, que armazena as informações do modelo, e o raio.

Depois da criação do modelo SOM, é possível verificar os valores que os neurônios adotam ao se organizar como centroides no modelo por meio do componente *codes*. Entretanto, os valores são gerados no padrão de normalização que usamos no dataset, e por consequência, precisamos passá-los para uma função também fornecida pela biblioteca RSNNS, que desempenha uma tarefa de normalização reversa. Dessa forma, é possível fazer o processo contrário de normalização, passando como parâmetro dados normalizados.

A seguir, criamos datasets para armazenar os dados que obtivemos da análise. Os dados dos pontos de demanda, ou cidades para nosso contexto são armazenados no data frame *customer\_locations*, contendo um índice *id,* a latitude, longitude, população em sua forma original, sem passar pelo processo de normalização, e a localização, sendo esta o centroide a qual a cidade está atrelada. Já os centroides, ou armazéns, são armazenados no data frame *warehouse\_locations*, possuindo os seguintes dados: um índice *id;* a latitude e longitude que passaram pelo processo reverso de normalização; variável *dist\_to\_mean*, que representa a distância entre cada centroide à média dos mesmos; *cost\_per\_square\_meter,* 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 (/m²), segundo quartil com 1500/m², terceiro quartil com 1000/m², e valores acima com 500/m². O data frame também conta com a população somada de todas as cidades atreladas a cada centroide, o tamanho do armazém em metros quadrados determinado pelo número de habitantes, e o custo do armazém, sendo calculado multiplicando o custo por metro quadrado pelo seu tamanho.

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. Para isso, foi criada uma função que recebe como parâmetros uma cidade do *customer\_locations­* e um centroide do ­*warehouse\_locations* e calcula a distância entre eles utilizando a função haversine() da biblioteca pracma...............................

O modelo de programação é criado nesse momento, utilizando como variáveis x do tipo binário, representando se uma cidade “i” está atrelada a um centroide “j”, e a variável y também do tipo binário, representando se um centroide “j” será montado ou não. É definida uma função de maximização utilizando a soma dos custos de transporte e a soma do custo de instalação de cada centroide escolhido. São adicionadas duas restrições ao modelo, sendo que a primeira indica que cada cidade precisa estar atrelada a um centroide, e a segunda que se uma cidade está atrelada a um centroide, este centroide deve ser construído.

Solucionamos o modelo por meio do solucionador glpk da infraestrutura de otimização do R ( R Optimization Infraestructure – ROI ), e coletamos o resultado.

1. RESULTADO E DISCUSSÃO

Como resultado do algoritmo descrito anteriormente, conseguimos obter por meio de modelos de programação matemática as localizações ótimas para o problema de localização de intalações.

1. CONSIDERAÇÕES FINAIS

Clique aqui para digitar o texto

1. REFERÊNCIAS

Clique aqui para digitar o texto

Contatos: e-mail aluno e e-mail orientador