

**Auxílio à distribuição geográfica de recursos
utilizando mineração de dados e aprendizado de
máquina**

M. G. Oliveira

Technical Report - RT-INF_001-11 - Relatório Técnico
June - 2011 - Junho

The contents of this document are the sole responsibility of the authors.
O conteúdo do presente documento é de única responsabilidade dos autores.

**Instituto de Informática
Universidade Federal de Goiás**
www.inf.ufg.br

Auxílio à distribuição geográfica de recursos utilizando mineração de dados e aprendizado de máquina

Max Gontijo de OLIVEIRA *
max.gontijo.oliveira@gmail.com

Cedric Luiz de CARVALHO †
cedric@inf.ufg.br

Abstract. *This paper is intended to propose a solution to the problem of resource distribution in a predetermined geographic region. For this purpose, techniques and data mining approaches have been studied, order to obtain patterns of demand of the distribution problem you want solved. This type of problem is known in literature as a facility location model, where we seek to find the best locations for new facilities considering the estimated demand.*

Keywords: data mining, clustering algorithms, facility location

Resumo. *Este trabalho tem o objetivo de apresentar uma proposta de solução para o problema de distribuição de recursos em uma região geográfica pré-determinada. Para tal propósito, técnicas e abordagens de mineração de dados vem sendo estudadas, visando a obtenção de padrões de demanda do problema de distribuição que se deseja atacar. Esse tipo de problema é conhecido na literatura como um modelo facility location, onde busca-se encontrar as melhores localizações para a inserção de uma facilidade considerando a demanda.*

Palavras-Chave: mineração de dados, algoritmos de cluster, facility location

1 Introdução

Decidir sobre como realizar a distribuição de facilidades¹ em uma região geográfica limitada (uma cidade, por exemplo), é uma atividade estratégia para muitas organizações. Quando a disponibilização de um recurso envolve um custo muito alto ou o número de recursos disponíveis é limitado, esse problema fica ainda mais preocupante, pois uma distribuição mal realizada pode acarretar em prejuízo certo para a organização.

Uma distribuição mal realizada de recursos disponíveis, incluem (mas não se limitam) interesse maior de clientes locais à concorrência, dificuldades de atingir metas de qualidade de

*Mestrando em Ciência da Computação, INF-UFG

†Orientador

¹Nesse contexto, uma facilidade é um recurso que provê algum tipo de serviço ou produto para atender à uma demanda medida ou estimada que, geralmente, está próxima a esse recurso. De modo geral, facilidade e recurso terão o mesmo significado nesse trabalho.

serviço (prazo para atendimento) e baixa demanda para o recurso. Somente para ilustrar, pode-se tomar como exemplo de organizações que se deparam com esse tipo de decisão, redes de supermercados (que precisam decidir onde instalar uma nova filial), empresas que coordenam o serviço de divulgação por panfletagem (que precisam decidir como alocar as pessoas que irão distribuir panfletos).

Esses problemas de realizar a distribuição de recursos (ou facilidades) em uma região geográfica limitada é caracterizado na literatura como *facility location*[5].

1.1 Objetivos

De modo geral, esses problemas lidam com recursos que, quando distribuídos, sempre permanecerão no local da distribuição. Isso é o caso das organizações que tem que construir edificações para o fornecimento do serviço ou produto.

Entretanto, esse trabalho visa atacar problemas onde os recursos distribuídos irão se locomover para atender as demandas. Cooperativas de taxi, por exemplo, poderiam oferecer um atendimento mais rápido (e mais barato para a cooperativa) se realizasse a distribuição dos taxistas de forma bem planejada, para que cada demanda recebida pudesse ser mais facilmente atendida. Atendimentos médicos de emergência podem ser realizados cada vez mais rápidos se as ambulâncias estiverem distribuídas em locais estratégicos. É importante notar que nesse tipo de problema, a facilidade deverá estar preparada para atender a uma demanda local, de modo que chegar à um consumidor não afaste a facilidade demais dos outros consumidores.

Segundo Benati e Laporte[1], problemas de otimização modelados como *facility location* são NP-Difíceis, tornando inviável até mesmo a solução de instâncias relativamente pequenas. Dessa forma, esse trabalho visa buscar uma abordagem em inteligência artificial utilizando técnicas de mineração para conseguir retornar boas sugestões de distribuição de recursos de um determinado problema.

2 Material e Método

Este trabalho está sendo realizado de forma investigativa. O problema foi bem definido, podendo ser resumido em realizar uma distribuição geográfica de recursos de modo a otimizar a qualidade do serviço e minimizar o custo do mesmo para a organização. A partir daí, vem sendo realizado um estudo aprofundado em documentações diversas publicadas pela comunidade científica, em especial, a comunidade de pesquisa operacional, na busca de uma solução viável para esse problema.

Inicialmente, o foco da pesquisa foi nos conceitos fundamentais de mineração de dados bem como um estudo geral sobre as diversas técnicas de mineração. Esse estudo foi realizado para que pudesse fundamentar a escolha de melhores algoritmos para a análise de dados históricos do problema.

Em seguida, o foco da pesquisa foi em características e abordagens para solução de problemas do tipo *facility location*. Nesse estudo, constatou-se que, embora haja considerável crescimento de pesquisas nessa área na última década[5], ainda não existem abordagens consolidadas e incontestavelmente eficientes. Entretanto, foram identificados trabalhos interessantes que apresentaram soluções para alguns tipos de problemas do tipo *facility location* utilizando algoritmos de clusterização no processo de mineração de dados. Rosing[6] utilizou algoritmos de clusterização para resolver o problema MWP² de forma heurística. Bramel e Simchi-Levi[2],

²MWP[7](*multi-source Weber problem*) se refere ao problema de Weber, que consiste em distribuir recursos

Fisher e Jaikumar[3] e Klose e Wittmann[4] apresentaram abordagens para problemas de distribuição de veículos e rotas utilizando algoritmos de clusterização.

2.1 Implementação

Implementações em Java vêm sendo realizadas, para realizar testes e complementar o estudo desses algoritmos de clusterização. Para auxiliar nessa tarefa, a API e ferramentas do projeto WEKA³ vêm sendo utilizadas. Dentre outras diversas funcionalidades, o WEKA provê recursos diversos para mineração de dados utilizando algoritmos de clusterização, como o apresentado na Figura 1.

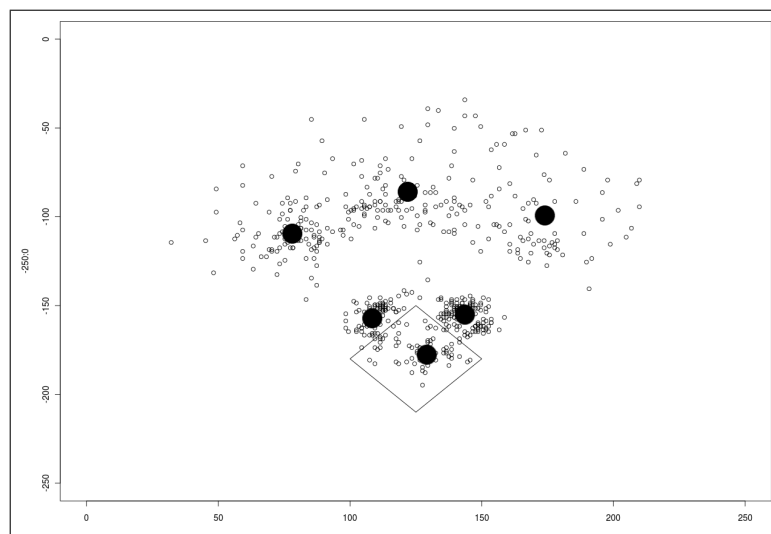


Figura 1: Ferramenta do WEKA de visualização de clusters identificados em uma mineração em dados históricos de uma base de dados qualquer. Os pontos pretos maiores denotam o epicentro de cada cluster.

Algoritmos de clusterização revelam agrupamentos de dados que, comumente, não são facilmente perceptíveis. Esses agrupamentos se dão pelas características mineradas.

Assim, seria uma abordagem razoavelmente plausível realizar agrupamentos por região, identificando assim, o epicentro geográfico dos clusters, e, conseqüentemente, a localização de cada viatura. Dessa forma, os estudos agora estão focados nesse ponto.

3 Resultados e discussões

Até o presente momento, a pesquisa está em andamento e ainda não foram gerados resultados apresentáveis. Entretanto, a identificação de abordagens e algoritmos de clusterização utilizados na resolução de problemas dessa natureza pode ser considerado um bom avanço.

Como estudo de caso, está sendo realizado paralelamente à pesquisa, um trabalho de busca de uma solução para a distribuição de viaturas da CELG⁴. Nesse problema real, o gerente de operações da CELG deve realizar sazonalmente, uma distribuição geral das viaturas

geograficamente e alocar demandas a esses recursos. Esse problema é NP-Difícil.

³Projeto que implementa diversos algoritmos de mineração de dados em JAVA.
<http://www.cs.waikato.ac.nz/ml/weka>

⁴Empresa distribuidora de energia elétrica no Estado de Goiás

disponíveis na região geográfica de cada central de operações, e deve fazer o mesmo em situações de emergência, quando o número de danos na rede elétrica aumenta muito e precisam ser atendidos rapidamente.

Assim, com esse problema sendo facilmente caracterizado como um problema *facility location* e com os estudos sobre esse assunto em andamento, atualmente vem sendo estudado com maior aprofundamento, algoritmos de clusterização e abordagens em que os mesmos são usados para resolver problemas dessa natureza.

A abordagem que está sendo estudada é a de utilizar atributos geográficos (dentre outros), como localização e distância, para identificar clusters que determinem agrupamentos geográficos que indiquem a posição de cada viatura. A Figura 2 mostra uma situação fictícia de um conjunto de ocorrências na cidade de Goiânia.

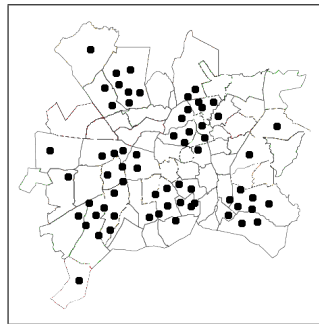


Figura 2: Mapa de Goiânia com ocorrências espalhadas

O que se deseja é que, ao aplicar algum algoritmo de clusterização (p-mediana, por exemplo), sejam identificadas os agrupamentos que possam indicar a localização de onde deveriam ser distribuídas as viaturas da CELG na região.

A Figura 3 ilustra bem isso, mostrando como seria a identificação dos seis agrupamentos quando a quantidade de recursos (viaturas) disponíveis é seis.

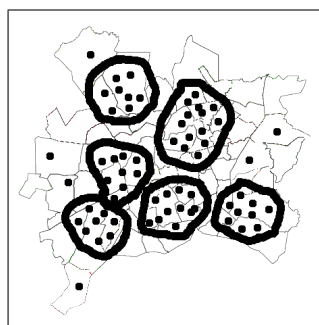


Figura 3: Mapa de Goiânia com ocorrências espalhadas e clusters identificados

4 Conclusão

Este trabalho tem o objetivo de propor uma solução para o problema de distribuição geográfica de recursos em ambientes cuja demanda é volátil.

Paralelamente, vem sendo desenvolvida uma solução específica para o problema de distribuição de viaturas da CELG, que devem ser acionadas quando ocorrer alguma queda no fornecimento de energia.

Algoritmos de aprendizado de máquina utilizando técnicas de mineração de dados vem sendo usados para a identificação das melhores abordagens e, caso necessário, uma abordagem adaptada pode vir a ser desenvolvida.

Por fim, essa distribuição assume que todos os recursos são de um mesmo tipo e que cada recurso é capaz de atender a toda e qualquer demanda do problema. Entretanto, em muitos casos, tais recursos podem ser de diferentes tipos, sendo a criação de uma abordagem que considere essas diferenças, um trabalho futuro interessante a ser iniciado.

Referências

- [1] Benati, S.; Laporte, G. **Tabu search algorithms for the (rlxp)-medianoid and (rlp)-centroid problems.** *Location Science* 2, 1994.
- [2] Bramel, J.; Simchi-Levi, D. **A location based heuristic for general routing problems.** *Operations Research*, 1995.
- [3] Fisher, M.; Jaikumar, R. **A generalized assignment heuristic for vehicle routing.** *Networks*.
- [4] Klose, A.; Wittmann, S. **Facility location based on clustering.** *Proceedings of the Second International Workshop on Distribution Logistics*, 1995.
- [5] Melo, M.; Nickel, S.; da Gama, F. S. **Facility location and supply chain management - A review.** *European Journal of Operational Research*, 2009.
- [6] Rosing, K. **The optimal location of steam generators in large heavy oil fields.** *American Journal of Mathematical and Management Sciences*, 1992.
- [7] Wesolowsky, G. **The Weber problem: History and perspectives.** *Location Science* 1, 1993.